

UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**CONSTRUÇÃO DE MODELOS PARA MÉDIAS E
VARIÂNCIAS NA OTIMIZAÇÃO EXPERIMENTAL
DE PRODUTOS E PROCESSOS**

Elaborado por **Pedro Alberto Barbetta** como requisito parcial
para obtenção do grau de Doutor em Engenharia de Produção.

Comissão Examinadora:

Prof. Robert Wayne Samohyl, Ph. D. - orientador

Prof. José Luis Duarte Ribeiro, Ph. D. - co-orientador (UFRGS)

Prof. João Riboldi, Dr. (UFRGS)

Prof. Adalberto Luiz Verani Depizzolatti, Dr. (UDESC)

Prof. Antonio Cezar Bornia, Dr.

Prof^ª. Sílvia Nassar, Dra. - moderadora

Florianópolis, 1998



0.296.617-5




UFSC-BU

UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
DE PRODUÇÃO

**CONSTRUÇÃO DE MODELOS PARA MÉDIAS E
VARIÂNCIAS NA OTIMIZAÇÃO EXPERIMENTAL DE
PRODUTOS E PROCESSOS**


Elaborado por
Pedro Alberto Barbeta

ESTA TESE FOI JULGADA ADEQUADA PARA A OBTENÇÃO DO
TÍTULO DE DOUTOR EM ENGENHARIA DE PRODUÇÃO, EM SUA
FORMA FINAL, PELO PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA DE PRODUÇÃO.

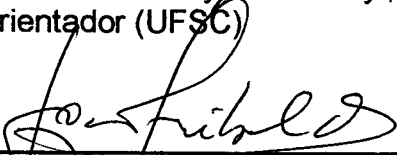


Prof. Ricardo Miranda Barcia, PhD.
Coordenador do Curso

BANCA EXAMINADORA



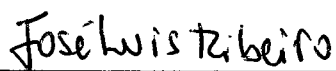
Prof. Robert Wayne Samohyl, PhD.
Orientador (UFSC)



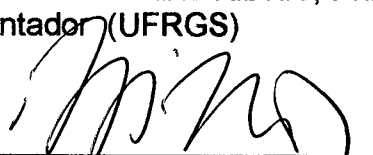
Prof. João Riboldi, Dr.
(UFRGS)



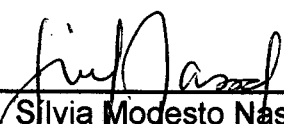
Prof. Antonio Cezar Bornia, Dr.
(UFSC)



Prof. José Luis Duarte Ribeiro, PhD.
Co-orientador (UFRGS)



Prof. Adalberto Luiz Verani Depizzolatti, Dr.
(UDESC)



Profª Silvia Modesto Nassar, Dra.
Moderadora (UFSC)

*À minha esposa Carmelita
e aos meus filhos Luiz e
Eduardo.*

AGRADECIMENTOS

Quero registrar os meus agradecimentos a todos que direta ou indiretamente contribuíram para a realização deste trabalho. Em especial,

à minha família, que foi bastante compreensiva, proporcionando um bom ambiente de trabalho;

aos colegas do Departamento de Informática e de Estatística da UFSC, que confiaram em mim e permitiram o meu afastamento para formação;

ao Programa de Pós-graduação em Engenharia de Produção da UFSC, que possibilitou a realização de um trabalho de Tese na área de meu interesse;

ao Prof. José L. D. Ribeiro da UFRGS, que na condição de co-orientador, dedicou bastante tempo na orientação deste trabalho, sempre apresentando os seus conhecimentos de forma bem clara e transmitindo bastante otimismo;

ao Prof. Robert W. Samohyl, orientador e amigo, que sempre incentivou o desenvolvimento do trabalho e ofereceu várias contribuições;

ao professor e amigo Antonio Cezar Bornia, pelas inúmeras sugestões que ofereceu no decorrer destes anos e pelas sábias correções de português na escrita do texto;

à professora e amiga Sílvia Nassar, pelo incentivo e pelas boas sugestões;

ao professor e amigo Adalberto L. V. Depizzolatti, pelas sugestões e correções na linguagem matemática e

ao Prof. João Riboldi da UFRGS, pelo pronto atendimento do convite em vir para Florianópolis participar do presente trabalho.

RESUMO

Nas duas últimas décadas, tem crescido bastante o interesse na otimização experimental, buscando-se melhoria da qualidade já na fase do projeto do produto ou processo. Neste enfoque, tenta-se encontrar uma combinação de níveis dos fatores do processo que garanta um produto com melhor qualidade, sem aumento nos custos. Isto pode ser feito a partir de experimentos estatisticamente planejados e da metodologia de superfície de resposta, onde normalmente é necessário ajustar funções matemáticas para o valor esperado e para a variância de características funcionais do produto, em função dos fatores controláveis do processo. Funções de variância são de grande importância quando se planeja produtos robustos a ruídos do processo, a variações de fatores ambientais e as diferentes condições de uso. Um procedimento bastante comum para se construir um modelo para a variância do processo baseia-se nas variâncias amostrais, calculadas sobre as replicações do experimento. Estas variâncias amostrais, após uma transformação logarítmica, são usadas como se fossem observações de um modelo de regressão. A transformação logarítmica faz com que a equação de regressão satisfaça aproximadamente os pressupostos básicos de um modelo linear, especialmente em experimentos com grande número de replicações. A presente pesquisa apresenta uma estratégia para se construir a equação de regressão para a variância em experimentos com poucas replicações, baseando-se em três procedimentos propostos: uma transformação logarítmica modificada, agregação de variâncias de fatores não significativos e combinação das variâncias amostrais com os resíduos do modelo de regressão da média. Os procedimentos são generalizados para situações de múltiplas respostas, onde se tem um sistema de equações para a média e um sistema de equações para a variância do processo. A eficiência dos procedimentos propostos são avaliados por simulações do tipo Monte Carlo e a estratégia de análise é ilustrada com três estudos de caso extraídos da literatura.

ABSTRACT

In the past two decades, the interest in experimental optimization has greatly increased, seeking an improvement in quality, even in the product or process design. From this perspective, an attempt is made to find a combination of levels of process factors that guarantee a product of the highest quality, with no increase in cost. This can be done from statistically planned experiments and the response surface methodology, in which it is usually necessary to adjust mathematical functions to the expected value and to the variance of the product's functional characteristics, in terms of the controllable factors of the process. Variance functions are of great importance in planning products that are robust to the noises of the process, to the variations in environmental factors and to the different conditions of use. A very common procedure for building a model for process variance is based on the sample variances, calculated from the replications of the experiment. These sample variances, after a logarithmic transformation, are used as though they were observations from a regression model. The logarithmic transformation causes the regression equation to satisfy approximately the basic premises of the linear model, especially in experiments with a large number of replications. This paper presents a strategy for building a regression equation for variance in experiments with few replications, based on three proposed procedures: a modified logarithmic transformation, aggregation of non-significant factor variances and a combination of sample variances with residuals of the mean regression model. The procedures are generalized for multiple response situations, in which there is a system of equations for the mean and a system of equations for the process variance. The efficiency of the procedures is evaluated by Monte Carlo-type simulations and the strategy of analysis is illustrated with three case studies extracted from the literature.

SUMÁRIO

1 - INTRODUÇÃO	1
1.1 - A qualidade em produtos e processos	1
1.2 - Tema da pesquisa	2
1.3 - Justificativa e relevância	7
1.4 - Objetivos	8
1.5 - Métodos de desenvolvimento da pesquisa	8
1.6 - Estrutura do trabalho	10
1.7 - Limitações da pesquisa	11
2 - REVISÃO DA LITERATURA	13
2.1 - Metodologia de superfície de resposta	13
2.1.1 - Planejamento e análise de experimentos	14
2.1.2 - Construção de modelos	19
2.1.3 - Análise da superfície de resposta e otimização	29
2.2 - Construção de modelos para a variância do processo	31
2.2.1 - Identificação dos efeitos de dispersão	34
2.2.2 - O modelo básico	36
2.2.3 - Modelos baseados em variâncias amostrais	38
2.2.4 - Modelos baseados nos resíduos	42
2.2.5 - Outros métodos de estimação	48
2.2.6 - Modelos com efeitos de locação e de dispersão	50
2.3 - Construção de modelos com múltiplas respostas	53
2.3.1 - Regressão multivariada	55
2.3.2 - Sistemas de equações de regressão	58
2.4 - Otimização e soluções de compromisso	64
2.4.1 - Uma característica de qualidade	65
2.4.2 - Múltiplas características de qualidade	68
3 - MODELOS DE VARIÂNCIA EM EXPERIMENTOS COM POUCAS REPLICAÇÕES	73
3.1 - Descrição geral das simulações	75
3.2 - O uso de fatores não significativos como pseudo-replicações	76

3.2.1 - Experimentos fatoriais 2^k não replicados	78
3.2.2 - Experimentos fatoriais 2^k com replicações	84
3.2.2 - Experimentos fatoriais 3^k	87
3.3 - A transformação logarítmica modificada	88
3.3.1 - A distribuição dos erros na equação de regressão proposta	90
3.3.2 - Fatores relevantes na eficiência da transformação proposta	93
3.3.3 - A transformação em experimentos fatoriais do tipo 2^k	100
3.3.4 - A transformação em experimentos do tipo 3^k e CCD	105
3.3.5 - Estudo do vício das estimativas	107
3.4 - Agregação das variâncias amostrais com os resíduos quadráticos médios	109
3.4.1 - O método da agregação	109
3.4.2 - Estudos de simulação	112
4 - ABORDAGEM MULTIVARIADA	116
4.1 - O modelo geral	116
4.2 - O sistema de equações das médias	117
4.2.1 - O modelo SUR com variâncias desiguais	118
4.2.2 - A suposição de correlações constantes	119
4.2.3 - Estimação dos parâmetros	122
4.3 - Sistema de equações das variâncias	124
4.4 - Algoritmo para a estimação conjunta dos parâmetros	127
4.5 - Estudos de simulação	129
4.6 - Robustez em relação à hipótese de correlações constantes	135
5 - ESTRATÉGIAS DE ANÁLISE E ESTUDOS DE CASO	137
5.1 - Estratégias de análise	137
5.1.1 - Experimentos não replicados	138
5.1.2 - Experimentos com replicações	140
5.1.3 - Construção dos modelos	142
5.1.4 - Experimentos com múltiplas respostas	144
5.2 - Estudo de caso nº 1: moldagem por injeção	145
5.2.1 - Identificação dos efeitos da média do processo	146
5.2.2 - Identificação dos efeitos da variância do processo	147

5.2.3 - Construção dos modelos da média e da variância	151
5.3 - Estudo de caso nº 2: o exemplo da catapulta	152
5.3.1 - Ajuste inicial do modelo da média	153
5.3.2 - Ajuste inicial do modelo da variância	156
5.3.3 - Estimação conjunta dos modelos da média e da variância	158
5.4 - Estudo de caso nº 3: formulação de uma massa cerâmica	161
5.4.1 - Modelos iniciais para as médias do processo	162
5.4.2 - Modelos para as variâncias do processo	165
5.4.3 - Modelos univariados para as médias usando MQG	167
5.4.4 - Estimação dos parâmetros dos modelos de forma sistêmica	168
6 - CONSIDERAÇÕES FINAIS	171
6.1 - Conclusões	171
6.2 - Sugestões para novas pesquisas	174
REFERÊNCIAS BIBLIOGRÁFICAS	176
APÊNDICE: PROGRAMAS COMPUTACIONAIS	185
A - Agregação de resíduos quadráticos: simulação	186
B - Agregação de variâncias amostrais: simulação	195
C - Distribuição $\log(\chi_1 + q)$	203
D - Avaliação da transformação logarítmica modificada: simulação	204
E - Agregação de variâncias amostrais e resíduos quadráticos: simulação	215
F - Correlações entre logaritmos de variâncias amostrais	220
G - MQG - SUR: simulação	222
H - Estudo de caso nº 1	230
I - Estudo de caso nº 2	233
J - Estudo de caso nº 3	238

LISTA DE TABELAS

Tabela 3.1 - Matriz experimental (2^{8-2}) e resultados da eficiência relativa (ef.) e do vício com a agregação (vic.) do estudo de simulação nº 1	81
Tabela 3.2 - Estimativa dos principais efeitos na eficiência relativa da agregação	83
Tabela 3.3 - Resultados da eficiência relativa com a agregação. Estudo de simulação nº 2	85
Tabela 3.4 - Resultados da eficiência relativa com a agregação em experimentos do tipo 3^3	87
Tabela 3.5 - Variâncias de $\log(Z_i^2 + c \bar{Z}^2)$, segundo a população geradora e valores de c	91
Tabela 3.6 - Algumas características da distribuição de $\log(Z_i^2 + c \bar{Z}^2)$	92
Tabela 3.7 - Resultados da eficiência relativa da transformação logarítmica modificada no estudo de simulação nº 3	95
Tabela 3.8 - Estimativa dos efeitos significativos do estudo de simulação nº 3	96
Tabela 3.9 - Projeto experimental da simulação nº 4 e os valores de c que maximizam a eficiência da transformação logarítmica modificada	102
Tabela 3.10 - Análise de variância do ajuste de um polinômio de segundo grau para prever o valor ótimo de c , após transformação logarítmica	102
Tabela 3.11 - Projeto experimental da simulação nº 5 e os valores de c que maximizam a eficiência da transformação logarítmica modificada	106
Tabela 3.12 - Coeficiente de variação e vício relativo dos estimadores com a transformação proposta	108
Tabela 4.1 - Valores de correlação entre o logaritmo de duas variâncias amostrais	126
Tabela 5.1 - Projeto experimental e os dados (grau de encolhimento) do experimento de moldagem por injeção	146
Tabela 5.2 - Resíduos do modelo da média e o esquema de cálculo da estatística F^* para todos os efeitos possíveis	148
Tabela 5.3 - Resíduos do modelo da média e transformações para a análise de identificação dos efeitos de dispersão	149
Tabela 5.4 - Fatores controláveis do processo usados no estudo experimental	152
Tabela 5.5 - Projeto experimental e resultados do experimento da catapulta	153
Tabela 5.6 - Construção de um modelo para a média do processo da catapulta: ajuste de um modelo quadrático completo por MQO	154
Tabela 5.7 - Avaliação do modelo especificado: verificação de falta de ajuste	154

Tabela 5.8 - Avaliação do modelo especificado: decomposição	154
Tabela 5.9 - Resíduos quadráticos médios calculados sobre os níveis do ruído	156
Tabela 5.10 - Construção de um modelo para a variância do processo da catapulta: ajuste por MQO de um modelo quadrático completo com transformação logarítmica modificada	157
Tabela 5.11 - Modelo final para a média do processo da catapulta	158
Tabela 5.12 - Modelo final para a variância do processo da catapulta. Estatísticas em termos do modelo exponencial	159
Tabela 5.13 - Comparação de procedimentos para estabelecer modelos para a variância, considerando os dados do experimento da catapulta	160
Tabela 5.14 - Níveis dos fatores controláveis do processo da massa cerâmica	162
Tabela 5.15 - Efeitos identificados em cada equação de regressão da média, coeficientes estimados por MQO, erros padrão das estimativas e algumas estatísticas de ajuste dos modelos	164
Tabela 5.16 - Efeitos identificados em cada equação de regressão da variância, coeficientes estimados por MQO com transformação $\log\{S_i^2 + (0,02)\bar{S}^2\}$, erros padrão das estimativas e algumas estatísticas de ajuste dos modelos	166
Tabela 5.17 - Equações de regressão das médias, com estimativas por MQG, tratando cada resposta isoladamente	167
Tabela 5.18 - Coeficientes de correlação entre os resíduos dos modelos das médias e entre os resíduos dos modelos das variâncias	168
Tabela 5.19 - Sistema de equações de regressão das médias, com estimativas obtidas pelo método MQG-SUR. Resultados após duas iterações	169

LISTA DE QUADROS

Quadro 3.1 - Fatores e níveis usados no estudo de simulação nº 1	78
Quadro 3.2 - Fatores e níveis usados no estudo de simulação nº 2	84
Quadro 3.3 - Fatores e níveis usados no estudo de simulação nº 3	93
Quadro 3.4 - Fatores e níveis usados no estudo de simulação nº 4	100
Quadro 3.5 - Fatores e níveis usados no estudo de simulação nº 5	106

LISTA DE FIGURAS

Fig. 1.1 - Ilustração de uma superfície de resposta e o gráfico de contornos dos fatores 1 e 2	4
Fig. 2.1 - Ilustração de um CCD com dois fatores e níveis codificados na forma padrão	19
Fig. 2.2 - Ilustração de funções perda quadrática em função do tipo de característica de qualidade	32
Fig. 2.3 - As principais características da distribuição do termo de erro de $\log(S^2)$, em função do número de replicações	40
Fig. 2.4 - Eficiência de $\log(S^2)$ em relação a S^2 na estimação do parâmetro σ^2 da distribuição normal. Dados extraídos de Bartlett e Kendall (1946)	41
Fig. 2.5 - Ilustração da densidade de probabilidade dos termos de erro de um par de medidas	60
Fig. 2.6 - Diferentes transformadas para a construção da função utilidade, considerando $s = t = 1$	71
Fig. 3.1 - Ilustração do processo de agregação de variâncias amostrais num experimento fatorial 2^3	77
Fig. 3.2 - Participação percentual da magnitude de cada efeito em relação à soma do valor absoluto de todos os efeitos (principais e interações de segunda ordem)	81
Fig. 3.3 - Gráfico normal dos efeitos (principais e interações de segunda ordem).	82
Fig. 3.4 - Análise dos resíduos de um modelo ajustado com os sete maiores efeitos da simulação nº 1	82
Fig. 3.5 - Gráfico normal dos efeitos dos dados da simulação nº 2, método RQ	86
Fig. 3.6 - A função logarítmica e a base das distribuições de S^2 e de $S^2 + q$	89
Fig. 3.7 - Diagramas de caixas de 10.000 observações simuladas da distribuição de $\log(Z_i^2 + c\bar{Z}^2)$, com $c = 0, 0,01$ e $0,10$	92
Fig. 3.8 - Gráfico normal dos efeitos do estudo de simulação nº 3	96
Fig. 3.9 - Eficiência da transformação proposta em função da percentagem da variância global adicionada às estatísticas. Experimentos com duas replicações	97
Fig. 3.10 - Eficiência da transformação proposta em função da percentagem da variância global adicionada às estatísticas. Experimentos com quatro replicações	98
Fig. 3.11 - Eficiência da transformação proposta em função de c e do parâmetro do modelo da variância em três tipos de experimentos	99
Fig. 3.12 - Análise dos resíduos do ajuste do modelo quadrático completo	103

Fig. 3.13 - Eficiência da transformação logarítmica modificada em função de c e valores preditos de c pela equação (3.21)	105
Fig. 3.14 - Comparação dos métodos S2, RQ e S2RQ em função de δ . Modelos especificados pela transformação logarítmica modificada	113
Fig. 3.15 - Comparação entre os métodos S2, RQ e S2RQ em quatro projetos experimentais	115
Fig. 4.1 - Relação entre a correlação das variâncias e a correlação das respostas	126
Fig. 4.2 - Eficiência dos métodos MQG, SUR e MQG-SUR em relação ao MQO. Experimento 2^4 , $\rho_{12} = 0,40$, primeira equação	131
Fig. 4.3 - Eficiência dos métodos MQG, SUR e MQG-SUR em relação ao MQO. Experimento 2^4 , $\rho_{12} = 0,40$, segunda equação	132
Fig. 4.4 - Eficiência dos métodos MQG, SUR e MQG-SUR em relação ao MQO. Experimento 2^4 , $\rho_{12} = 0,80$, primeira equação	133
Fig. 4.5 - Eficiência dos métodos MQG, SUR e MQG-SUR em relação ao MQO. Experimento 2^4 , $\rho_{12} = 0,80$, segunda equação	133
Fig. 4.6 - Eficiência dos métodos MQG, SUR e MQG-SUR em relação ao MQO. Experimento CCD, $\rho_{12} = 0,40$, primeira equação	134
Fig. 4.7 - Eficiência dos métodos MQG, SUR e MQG-SUR em relação ao MQO. Experimento CCD, $\rho_{12} = 0,40$, segunda equação	134
Fig. 4.8 - Eficiência dos métodos MQG, SUR e MQG-SUR em relação ao MQO. Experimento CCD, $\rho_{12} = 0,80$, primeira equação	134
Fig. 4.9 - Eficiência dos métodos MQG, SUR e MQG-SUR em relação ao MQO. Experimento CCD, $\rho_{12} = 0,80$, segunda equação	135
Fig. 4.10 - Eficiência dos métodos MQG e MQG-SUR em relação ao MQO. Dados gerados de duas formas: (a) ρ_{12} fixo em 0,60 e (b) ρ_{12} variando de 0,40 a 0,80	136
Fig. 5.1 - Estratégia geral para construir modelos para a média e para a variância da resposta	143
Fig. 5.2 - Estratégias para a construção de modelos com múltiplas respostas	145
Fig. 5.1 - Gráfico normal dos efeitos da média do processo	147
Fig. 5.2 - Análise dos resíduos do modelo da média	147
Fig. 5.3 - Gráfico normal dos efeitos de dispersão - estatística F_i^* de Box e Myers	148
Fig. 5.4 - Gráfico normal dos efeitos de dispersão - estatística $\log(\hat{\epsilon}_i^2)$	149
Fig. 5.5 - Gráfico normal dos efeitos de dispersão - estatística $\log(\hat{\epsilon}_i^2 + \bar{\epsilon}^2)$	150
Fig. 5.6 - Esboço da catapulta com indicação dos fatores estudados no experimento	152

Fig. 5.7 -	Análise de resíduos do modelo da média. Estimação por MQO	155
Fig. 5.8 -	Análise de resíduos do modelo da variância. Estimação por MQO após transformação logarítmica modificada	157
Fig. 5.9 -	Análise de resíduos do modelo final da média. Estimação por MQG	158
Fig. 5.10 -	Análise de resíduos do modelo final da variância, construído com os resíduos do modelo da média, o qual foi ajustado por MQG	159
Fig. 5.11 -	Gráficos dos resíduos dos três modelos das médias ajustados por MQO	164
Fig. 5.12 -	Análise dos resíduos dos três modelos das variâncias, ajustados por MQO, após transformação $\log\{S_i^2 + (0,02)\bar{S}^2\}$	166
Fig. 5.13 -	Análise dos resíduos do modelo da média da retração linear, ajustado por MQG	168

LISTA DOS PRINCIPAIS SÍMBOLOS E SIGLAS

COM LETRAS LATINAS:

c	constante usada na transformação logarítmica modificada (seção 3.3)
\hat{c}	predição da constante usada na transformação logarítmica modificada (seções 3.3.3 e 3.3.4)
CCD	projeto de experimento do tipo central composto (seção 2.1.1)
Cov	covariância (seção 4.3)
Corr	correlação (seção 2.3.2)
E	valor esperado (seção 2.1.2)
\hat{e}_i	resíduo referente à i -ésima observação (seção 2.2.4)
\hat{e}_i^2	resíduo quadrático referente à i -ésima observação (seção 2.2.4) / resíduo quadrático médio relativo ao i -ésimo ponto experimental (cap. 3)
$\bar{\hat{e}}^2$	média dos resíduos quadráticos ou estimativa do erro quadrático médio (seção 3.3)
F_0	estatística de um teste com distribuição F , pode aparecer com os graus de liberdade entre parêntesis (seção 2.1.2)
F^*	estatística de Box e Myers, $= F_i^*$ (seção 2.2.1)
L	função perda quadrática (seção 1.2)
M	número de pontos experimentais ou número de tratamentos (seção 2.2.2)
m	número de replicações em cada ponto experimental (seção 2.2.2)
MQO	método dos mínimos quadrados ordinários (seção 2.1.2)
MQG	método dos mínimos quadrados generalizados (seção 2.2.4)
MQG - SUR	método que agrega sistema SUR e MQG (seção 4.2.3)
n	número de observações, $= Mm$ (seção 2.1.2)
p_F	nível descritivo de significância baseado num teste F (seção 3.4)
RQ	métodos baseados nos resíduos quadráticos (seção 2.2.4)
R^2	coeficiente de determinação (seção 2.1.2)
R_{aj}^2	coeficiente de determinação ajustado pelos graus de liberdade (seção 2.1.2)
S_2	métodos baseados nas variâncias amostrais (seção 2.2.3)
S2RQ	método que agrega variâncias amostrais e resíduos quadráticos (seção 3.4)
S^2	variância amostral (seção 2.3.2) / matriz de variâncias e covariâncias amostrais (seção 2.3.1)
S_i^2	variância amostral no i -ésimo ponto experimental (seção 2.2.3)

\overline{S}^2	média das variâncias amostrais (seção 3.3)
$s_{\ell\ell}$	variância amostral do termo de erro da ℓ -ésima resposta (seção 2.3.2)
$s_{\ell k}$	covariância amostral entre os termos de erro da ℓ -ésima e da k -ésima resposta (seção 2.3.2)
SUR	sistema de equações de regressão aparentemente não relacionadas (seção 2.3.2)
t_0	estatística de um teste com distribuição t (seção 2.1.2)
Var	variância (seção 2.2.1) / matriz de variâncias e covariâncias (seção 4.2.2)
v_i	medidas de variância que agregam variâncias amostrais e resíduos quadráticos (seção 3.4)
\bar{v}	média aritmética dos v_i ($i = 1, \dots, M$) (seção 3.4)
$x = (x_1, \dots, x_k)'$	vetor dos fatores controláveis do processo (seção 2.1)
X	matriz de planejamento do modelo da média (seção 2.1.2) / matriz de planejamento do sistema de médias (seção 2.3.2)
y	vetor das observações da resposta (seção 2.1.2) / vetor das observações de múltiplas respostas (seção 2.3.2)
Y	matriz de observações multivariadas (seção 2.3.1)
Z	matriz de planejamento do modelo da variância (seção 2.1.2)

COM LETRAS GREGAS:

β	vetor de parâmetros do modelo da média (seção 2.1.2) / matriz de parâmetros do modelo multivariado da média (seção 2.3.1) / vetor de parâmetros do sistema das médias (2.3.2)
$\hat{\beta}$	estimador ou estimativa de β (seções 2.1.2, 2.3.1 e 2.3.2)
θ	vetor de parâmetros do modelo da variância (seção 2.2.2)
$\hat{\theta}$	estimador ou estimativa de θ (seção 2.2)
μ_y	valor esperado de y , $= \mu$ (seção 2.2)
$\hat{\mu}_y$	estimador ou estimativa do valor esperado de y , $= \hat{\mu}$ (seção 2.2.4)
ρ	matriz de correlações (seção 4.2.2)
$\hat{\rho}$	estimador ou estimativa da matriz de correlações (seção 4.2.2)
$\rho_{k\ell}$	correlação entre a k -ésima e a ℓ -ésima resposta (seção 4.2.2)
$\hat{\rho}_{k\ell}$	estimador ou estimativa da correlação entre a k -ésima e a ℓ -ésima resposta (seção 4.2.2)
σ_y^2	variância de y , $= \sigma^2$ (seção 2.2)

$\hat{\sigma}_y^2$	estimador ou estimativa da variância de y , $= \hat{\sigma}^2$ (seção 2.2.4)
Σ	matriz de variâncias e covariâncias das respostas (seção 2.3.1)
Σ^*	matriz de variâncias e covariâncias do termo de erro do sistema SUR (seção 2.3.2)

OUTROS:

\otimes	produto de Kroneker (seção 2.3.2)
2^k , 2^{k-r} e 3^k	tipos de projetos de experimentos (seção 2.1.1)

INTRODUÇÃO

1.1 - A QUALIDADE EM PRODUTOS E PROCESSOS

Com o advento da globalização e a conseqüente competição acirrada para alcançar os consumidores, a busca por melhoria contínua da qualidade torna-se uma questão de sobrevivência para empresas e organizações.

A qualidade é adequação ao uso (Juran e Gryna, 1991, p. 21). Ou seja, um produto (bem tangível, serviço ou método) tem melhor qualidade se ele é mais adequado ao usuário a que se destina. Neste enfoque, a atenção ao cliente torna-se fundamental para se ter alta qualidade, pois um processo só tem melhoria efetiva de qualidade se ele gera benefícios concretos ao cliente.

Tendo a satisfação do cliente como objetivo principal, a qualidade deve ser projetada e desenvolvida no processo produtivo. Dos clientes, levantam-se as características que eles desejam, enquanto no processo produtivo se criam as condições para que o produto atenda aos anseios do consumidor, inclusive em termos de adequação do preço. Conforme argumenta Taguchi et al. (1990), o preço representa para o consumidor uma perda na hora da compra e a baixa qualidade representa uma perda adicional durante o uso do produto. Um dos objetivos centrais da engenharia da qualidade é a redução da perda total para o consumidor.

No processo produtivo, a qualidade é construída não somente na produção, mas nas diversas etapas do sistema produtivo. De modo ideal, as atividades do sistema devem interagir para que os produtos tenham suas características funcionais com desvios mínimos dos valores fixados no projeto, pois estes desvios constituem num exemplo típico de perda. Mantendo-se as características funcionais suficientemente próximas dos valores projetados, tem-se uma redução da perda para o consumidor e uma diminuição dos custos da qualidade.

A qualidade começa a ser construída no planejamento do produto, onde deve-se escutar o cliente para adequar as características e as especificações do produto em estudo. No projeto do produto, pode-se desenvolver um protótipo básico que desempenha as funções desejadas e exigidas com mínima variação possível das especificações projetadas. Experimentos podem ser realizados para se encontrar os níveis dos fatores do sistema que minimizam os desvios das características funcionais em relação aos valores nominais do projeto. Estabelecem-se, desta forma, os níveis ótimos dos fatores e as tolerâncias aceitáveis.

Concluído o projeto do produto, busca-se determinar processos que possam fabricá-lo com menor custo, mas dentro dos limites e tolerâncias previamente especificadas. Nesta fase, recorre-se também a experimentos para determinar os níveis operacionais do processo (temperatura, pressão, etc.), que levam as características funcionais do produto tão mais próximas possível dos valores projetados.

Na produção, a qualidade deve ser constantemente monitorada. Avaliações devem ser feitas nos pontos críticos do processo para estabelecer padrões e, depois, verificar se o processo mantém-se nos padrões estabelecidos. Recursos estatísticos, tais como os gráficos de controle, são de grande valia nesta fase. Ressalta-se a relevância do comprometimento de todos os funcionários com a qualidade para que o monitoramento não resulte apenas num controle estatístico, mas num instrumento de aprimoramento da qualidade.

A assistência técnica também deve ser considerada relevante num programa de melhoria da qualidade, pois através dela pode-se corrigir eventuais falhas do projeto ou da produção. Além disso, tem-se uma retroalimentação, permitindo reformular o projeto do produto ou planejar novos produtos.

1.2 - TEMA DA PESQUISA

A presente pesquisa procura oferecer uma contribuição na área da engenharia da qualidade. Mais especificamente, nos problemas de otimização experimental de produtos e processos, onde procura-se projetar produtos e processos com melhor qualidade, a partir dos resultados de experimentos estatisticamente planejados.

Nesta área, é comum avaliar a qualidade por uma função perda, a qual imputa prejuízos financeiros a qualquer inadequação do produto ao consumidor. Operacionalmente, considera-se que a perda aumenta na medida em que características funcionais do produto afastam-se de seus valores ideais. Por exemplo, quando uma máquina de produzir porcas de parafusos estiver programada para produzir as porcas com 9,00 mm de diâmetro, mas tender a produzi-las com 9,01 mm, estará havendo perda, mesmo que este valor esteja dentro do intervalo de especificação. A perda também aumenta quando houver aumento de variação entre os itens, ainda que, em média, o valor do diâmetro esteja no alvo. Neste sentido, obtêm-se melhores níveis de qualidade na medida em que o processo esteja ajustado aos seus valores ideais e, ao mesmo tempo, a variância devida a fatores perturbadores do processo, a fatores ambientais ou a outros ruídos, seja a mínima possível.

Na engenharia da qualidade, procura-se obter melhoria contínua de produtos e processos sem aumentar os custos de produção. Isto pode ser alcançado com a realização de estudos experimentais que permitam encontrar os níveis adequados dos fatores controláveis do processo, no sentido de ajustar o processo ao alvo e minimizar a sua variância. É o chamado *projeto de parâmetros* ou *projeto robusto*, enfatizado pelo engenheiro japonês Genichi Taguchi e introduzido no mundo ocidental a partir da década de 80. Fiod Neto (1997) ressalta que, em lugar de remover os ruídos, o que pode ser de difícil execução ou pode encarecer o processo produtivo, os métodos de Taguchi ambicionam remover os efeitos negativos gerados pelos ruídos, sem aumentar os custos de produção. Oh (1995) ressalta que, tradicionalmente, na análise de um projeto, a variabilidade era monitorada, taxada e controlada. Na nova abordagem, procura-se antecipar e acomodar a variabilidade, tornando o projeto robusto a ela.

Uma maneira eficiente de desenvolver o projeto de parâmetros é através da *metodologia de superfície de resposta* (RSM). Esta metodologia é uma combinação de técnicas de planejamento de experimentos, análise de regressão e métodos de otimização. A figura 1.1 ilustra uma superfície de resposta, onde se busca o máximo de uma *resposta* (característica funcional do produto) em função de dois fatores controláveis.

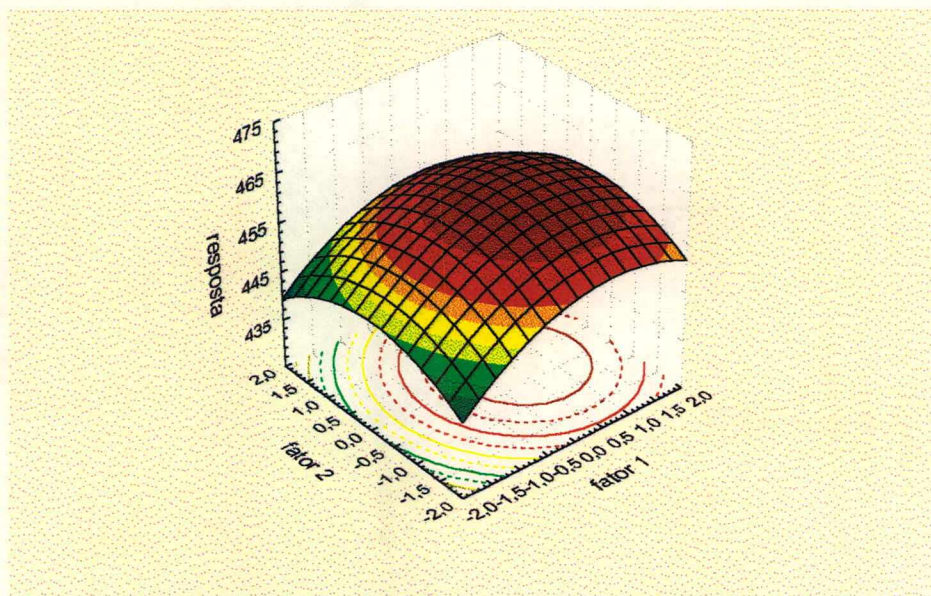


Fig. 1.1 - Ilustração de uma superfície de resposta e o gráfico de contornos dos fatores 1 e 2.

A superfície de resposta normalmente é uma equação de regressão, cujos parâmetros são estimados a partir de dados provenientes de um experimento estatisticamente planejado. A variável dependente da equação de regressão pode ser uma característica funcional do produto ou alguma estatística associada a esta característica, tal como a média ou a variância das observações. A característica funcional do produto é entendida como uma característica de qualidade do produto ou processo. Para ilustrar melhor a RSM num processo de melhoria da qualidade, utiliza-se, a seguir, o problema de otimização de uma máquina de estampar, que foi discutido por vários autores, dentre eles Vining e Myers (1990), Coppeland e Nelson (1996) e Guedes (1996).

O problema consiste em otimizar a capacidade que uma máquina de estampar tem para colorir etiquetas de pacotes, considerando os fatores velocidade (x_1), pressão (x_2) e distância (x_3). O estudo utiliza um planejamento experimental, onde cada fator é ensaiado em três níveis, totalizando 27 pontos experimentais (combinações dos níveis de x_1 , x_2 e x_3). Em cada ponto experimental, realizam-se três ensaios e mede-se, em cada ensaio, a capacidade de impressão da máquina (y). Na terminologia de planejamento de experimentos, tem-se um planejamento fatorial completo 3^3 com 3 replicações. A característica y é chamada de resposta.

Com os dados do experimento, Vining e Myers obtiveram a seguinte equação de regressão para o valor esperado de y :

$$\hat{\mu}_y = 327,6 + 177,0x_1 + 109,4x_2 + 131,5x_3 + 32,0x_1^2 - 22,4x_2^2 - 29,1x_3^2 + 66,0x_1x_2 + 75,5x_1x_3 + 43,6x_2x_3 \quad (1.1)$$

Por simplicidade operacional, consideraram-se os níveis ensaiados de x_1 , x_2 e x_3 codificados como -1 , 0 e 1 , correspondendo ao menor, intermediário e maior nível, respectivamente. A superfície de resposta foi assumida como um polinômio de segundo grau, definido sobre a região experimental. E os coeficientes da equação 1.1 foram obtidos através do método dos mínimos quadrados aplicado sobre os dados experimentais.

O objetivo é encontrar os níveis de x_1 , x_2 e x_3 que maximizam a capacidade de impressão da máquina, avaliada por $\hat{\mu}_y$. Contudo, para se ter um processo robusto e, conseqüentemente, melhor qualidade, torna-se necessário que os níveis escolhidos de x_1 , x_2 e x_3 sejam tais que acarretem uma variabilidade para y tão pequena quanto possível. Esta variabilidade é devida a fatores de ruído que agem de forma aleatória sobre os ensaios de cada ponto experimental. Ressalta-se, assim, a relevância de se obter uma equação de regressão que associe variações de x_1 , x_2 e x_3 com uma medida de variabilidade de y .

Vining e Myers (1990) usaram o desvio padrão amostral, calculado em cada ponto experimental, como se fossem observações e, assim, obtiveram, pelo método dos mínimos quadrados, a seguinte equação de regressão para o desvio padrão de y :

$$\hat{\sigma}_y = 34,9 + 11,5x_1 + 15,3x_2 + 29,2x_3 + 4,2x_1^2 - 1,3x_2^2 + 16,8x_3^2 + 7,7x_1x_2 + 5,1x_1x_3 + 14,1x_2x_3 \quad (1.2)$$

Tem-se uma segunda superfície de resposta e busca-se uma solução de compromisso entre ajustar a média do processo ao alvo (no caso, maximizar $\hat{\mu}_y$) e, simultaneamente, reduzir a variabilidade (minimizar $\hat{\sigma}_y$). Existem várias propostas na literatura para se obter uma solução de compromisso com estes dois objetivos. Algumas delas serão discutidas neste trabalho.

A presente pesquisa enfoca o problema de construção das superfícies de resposta. A literatura é bastante abrangente na questão de se obter uma equação de regressão para o valor esperado da resposta. A técnica tradicional dos mínimos quadrados, aplicada sobre as observações, presta-se a este objetivo. Por outro lado, quando se quer estabelecer uma equação de regressão em termos da variabilidade da resposta, ainda não se tem uma teoria consagrada. E a modelagem da variabilidade do processo é fundamental no desenvolvimento de projeto robusto via RSM.

Embora a medida de variabilidade possa ser genérica, este trabalho enfatiza o uso da variância, σ_y^2 . Esta ênfase é devida à fórmula clássica de se avaliarem desvios de uma característica y em relação a um valor estabelecido, τ . Esta fórmula, que é conhecida como perda quadrática média, ou erro quadrático médio, é definida por¹

$$E\{L(y)\} = E\{(y - \tau)^2\} \quad (1.3)$$

onde: $L(y)$ é a perda ocasionada por um particular valor de y e

$E\{.\}$ representa o operador de valor esperado ou valor médio.

A expressão (1.3) pode ser decomposta como²

$$E\{(y - \tau)^2\} = (\mu_y - \tau)^2 + \sigma_y^2 \quad (1.4)$$

A equação 1.4 mostra que para se reduzir a perda quadrática média deve-se ajustar a média μ_y ao alvo τ e minimizar a variância σ_y^2 . Ou seja, (1.4) sintetiza as idéias básicas do projeto de parâmetros, discutidas anteriormente.

Na prática, um produto ou processo costuma ter múltiplas características de qualidade. Por exemplo, ao avaliar a qualidade de um material, pode ser relevante a sua resistência, elasticidade, porosidade, etc. Este trabalho também contempla a construção de superfícies de resposta para múltiplas características de qualidade, utilizando-se técnicas de regressão multivariada e de sistemas de equações de regressão. A abordagem multivariada mantém a preocupação de se construírem

¹ Na expressão da perda, $L(y)$, costuma-se incluir um fator k que transforma os desvios, $y - \tau$, em perdas monetárias. Mas, no presente contexto, este fator é irrelevante.

² $E\{(y - \tau)^2\} = E\{(y - \mu_y + \mu_y - \tau)^2\} = E\{(\mu_y - \tau)^2\} + E\{(y - \mu_y)^2\} + 2E\{(y - \mu_y)(\mu_y - \tau)\} = (\mu_y - \tau)^2 + \sigma_y^2$, pois o último termo se anula: $E\{(y - \mu_y)(\mu_y - \tau)\} = (\mu_y - \tau) E\{(y - \mu_y)\} = (\mu_y - \tau)(\mu_y - \mu_y) = 0$.

superfícies de resposta tanto para as médias (sistema de equações para as médias) quanto para as variâncias (sistema de equações para as variâncias).

1.3 - JUSTIFICATIVA E RELEVÂNCIA

Nos últimos anos, a busca por melhoria contínua da qualidade tem sido vital para a sobrevivência das organizações, especialmente com o advento da globalização e o conseqüente acirramento da competitividade. Na década de 80, Taguchi e seus seguidores popularizaram a utilização do planejamento de experimento dentro de uma metodologia para a melhoria da qualidade. Esta abordagem é normalmente referenciada como *qualidade fora da linha*, por ser realizada na fase do planejamento do produto ou processo. O sucesso do uso de planejamento de experimentos na melhoria da qualidade é enfatizada por vários autores, dentre eles Box et al. (1978), Taguchi (1987), Phadke (1989), Tsui (1992), Lucas (1994), Myers e Montgomery (1995) e Fiod Neto (1997).

Existe, porém, bastante polêmica sobre as técnicas de planejamento de experimentos e formas de análises dos resultados. O presente trabalho segue a linha clássica de planejamento de experimentos e metodologia de superfície de resposta (RSM) por serem consideradas teoricamente mais consistentes e eficientes, conforme atestam Box (1988), Nair (1992), Lucas (1994), Myers e Montgomery (1995) e muitos outros. A outra abordagem é referenciada como Métodos de Taguchi, como apresentada em Phadke (1989).

Para se realizar a RSM é fundamental se ter boas estimativas para os parâmetros dos modelos de regressão, donde justificam-se técnicas que melhorem o processo de estimação. Além disso, pela abrangência de aplicações da RSM, o investimento neste processo torna-se bastante relevante.

Cabe observar que, numa aplicação particular, o sucesso de um projeto experimental na engenharia da qualidade depende muito do bom conhecimento do problema, da interação entre engenheiros que conhecem o processo e os profissionais que planejam os experimentos, da execução adequada dos experimentos e da correta interpretação dos resultados estatísticos. A escolha de bons métodos de estimação corresponde a uma pequena parte neste tipo de estudo. Contudo, estes métodos

geralmente não dependem da particular aplicação e, como consequência, pequenas melhorias sobre eles elevam a qualidade de inúmeras pesquisas aplicadas, resultando num ganho global bastante grande para a área.

1.4 - OBJETIVOS

O objetivo geral desta pesquisa é obter formas eficientes e relativamente simples de se estabelecerem equações de regressão para a média e para a variância de características de qualidade, com vistas à otimização experimental de produtos e processos. Mais especificamente, procura-se atingir os seguintes objetivos:

- 1) revisar os principais conceitos e técnicas associadas à metodologia de superfície de resposta e como esta metodologia pode incorporar os conceitos de engenharia de Taguchi;
- 2) levantar os métodos mais usados para se estabelecerem equações de regressão para a variância do processo;
- 3) propor métodos alternativos para se estimarem os parâmetros de modelos para a variância, que sejam simples, eficientes e, na medida do possível, robustos quanto à má especificação dos modelos;
- 4) adaptar procedimentos estatísticos de estimação multivariada para se obterem equações de regressão de médias e de variâncias relativas a várias características de qualidade do produto ou processo;
- 5) verificar a adequação dos métodos propostos e as condições em que eles apresentam melhores desempenhos e
- 6) aplicar os métodos propostos em alguns estudos de caso.

1.5 - MÉTODOS DE DESENVOLVIMENTO DA PESQUISA

Para se alcançar os dois primeiros objetivos, desenvolve-se um estudo na literatura de projeto de parâmetros, metodologia de superfície de resposta e modelos de regressão. Procura-se chegar ao estado da arte em modelos de regressão para a variância. O estudo aprofundado destes temas podem permitir desenvolver novas técnicas, alcançando-se, assim, o terceiro objetivo. Ao se incluir na revisão da

literatura métodos estatísticos multivariados e sistema de equações de regressão, tenta-se alcançar o quarto objetivo.

A condução e as conclusões desta pesquisa baseiam-se na revisão da literatura, em simulações do tipo Monte Carlo e na própria metodologia de planejamento de experimentos. Segundo Law e Kelton (1982), se as relações que compõem o modelo em estudo são suficientemente simples, torna-se possível o uso de métodos matemáticos para obter as informações exatas sobre as questões de interesse (soluções analíticas). Contudo, em situações mais complexas, modelos realistas são muito difíceis de serem tratados analiticamente, donde o apoio de métodos de simulação torna-se bastante útil. Numa simulação, utilizam-se recursos computacionais para, a partir de dados gerados por um modelo, se avaliarem numericamente as suas relações. As simulações do tipo Monte Carlo caracterizam-se por envolverem um esquema que emprega números aleatórios e são usadas para resolver problemas probabilísticos ou determinísticos, mas geralmente de natureza estática.

Na presente pesquisa, descrevem-se os modelos de acordo com situações típicas de problemas de projeto de parâmetros. Os modelos usados nas simulações são equações que relacionam a média e a variância de características de qualidade com fatores do processo. E a avaliação das técnicas propostas são feitas através da comparação dos valores estimados com os valores estabelecidos nos modelos da simulação.

Empregam-se simulações Monte Carlo, nesta pesquisa, sob dois enfoques. O primeiro caracteriza-se por análises exploratórias, onde os resultados da simulação permitem sinalizar os caminhos a seguir. Sob este enfoque, estudam-se possíveis combinações de técnicas descritas na literatura. Quando os resultados da simulação apontam um desempenho ruim do método, muda-se o rumo das investigações. Por outro lado, quando os resultados são promissores, considera-se que se alcançou parte do terceiro objetivo e investigam-se as técnicas com mais detalhes.

Num segundo enfoque, utilizam-se simulações Monte Carlo sob um projeto experimental, para se verificar a adequação das técnicas propostas e as condições em que elas têm melhor desempenho (objetivo cinco do trabalho). Nesta fase, tem-se

uma pesquisa do tipo experimental, onde provam-se relações de eficiência entre as técnicas analisadas sob as condições da realização do experimento. Em geral, pequenos estudos complementares de simulação são desenvolvidos para confirmar certas inferências ou, ainda, para tentar ampliar as conclusões.

Para contemplar o objetivo 6, buscam-se exemplos na literatura. A pesquisa inclui textos didáticos, artigos em revistas especializadas, teses e relatórios técnicos. Procuram-se estudos de caso que contemplem as diversas situações em que se propõem técnicas alternativas.

1.6 - ESTRUTURA DO TRABALHO

O capítulo 2 apresenta uma revisão da literatura, a qual engloba as idéias básicas da RSM. Procura-se chegar ao estado da arte na questão da estimação de parâmetros em modelos de regressão que relacionam variâncias com fatores controláveis do processo. Discute-se, também, como inserir este tipo de equação dentro da RSM, considerando o objetivo de minimizar a variância e levar a média do processo ao valor especificado. Este capítulo também apresenta um resumo de regressão multivariada e sistemas de equações de regressão aparentemente não relacionadas (SUR), como formas de generalizar a discussão para os casos em que o produto ou processo é avaliado por múltiplas características de qualidade.

O capítulo 3 centraliza a discussão no problema de se estimarem os parâmetros de um modelo de variância. É neste capítulo que se propõem formas alternativas de se estabelecer a equação de regressão da variância, com vistas a obter melhores estimativas pelo método dos mínimos quadrados. A verificação da eficiência das novas técnicas, via simulação Monte Carlo, também é realizada neste capítulo.

O capítulo 4 discute o problema de estimação dos modelos de regressão quando se tem múltiplas características de qualidade. É proposto um algoritmo para se obterem estimativas eficientes, baseando-se na teoria de sistemas de equações de regressão e nas técnicas discutidas no capítulo 3. A verificação da eficiência deste algoritmo também é feita através de simulações do tipo Monte Carlo.

O capítulo 5 apresenta uma estratégia de análise de dados experimentais com vistas a construir modelos para a média e para a variância do processo, fazendo-se uso dos procedimentos discutidos nos capítulos 3 e 4. Esta estratégia é ilustrada com alguns estudos de caso extraídos da literatura.

O capítulo 6 apresenta as considerações finais e propostas de novas investigações científicas. O trabalho completa-se com os programas computacionais utilizados nos capítulos 3, 4 e 5, apresentados no apêndice.

1.7 - LIMITAÇÕES DA PESQUISA

O trabalho pode ser visto como um conjunto de técnicas para resolver problemas de projeto de parâmetros via RSM. Contudo, a ênfase da pesquisa concentra-se em alguns aspectos estatísticos, deixando para um segundo plano diversos aspectos relacionados com a engenharia do processo, fundamentais para o sucesso do projeto de parâmetros.

O projeto de parâmetros ou projeto robusto é uma parte da engenharia da qualidade ou mesmo da otimização experimental de produtos e processos. No projeto de parâmetros, considera-se que já se tenha definido um produto ou processo a ser otimizado. Além disso, supõe-se que a melhoria possa ser obtida por escolhas adequadas dos níveis dos fatores controláveis do processo. Não se inclui aqui uma análise dos custos dos diferentes níveis dos fatores controláveis. Porém considera-se que, melhorando a qualidade do produto ou processo, estão se reduzindo os custos devido à má qualidade.

A construção das superfícies de resposta, que constitui parte importante da metodologia adotada neste trabalho, envolve a identificação dos fatores relevantes e a estimação de seus coeficientes. A pesquisa procura oferecer contribuição original na questão da estimação, concentrando-se o estudo no método de estimação conhecido como de mínimos quadrados. Não faz parte do escopo deste trabalho a comparação das técnicas propostas sob o enfoque de outros métodos de estimação, como o método da máxima verossimilhança, por exemplo. O trabalho também não se preocupa em verificar as vantagens (ou desvantagens) das técnicas propostas como

instrumentos de identificação dos efeitos significativos, ou seja, sob o enfoque de testes de hipóteses.

Outra limitação refere-se ao método de investigação de Monte Carlo, pois, mesmo que este seja realizado sob experimentos estatisticamente planejados, as suas conclusões são restritas ao conjunto de modelos ensaiados ou a combinações destes. Também não se buscam soluções analíticas, pois os modelos matemáticos e estatísticos deste trabalho são demasiadamente complexos.

O trabalho também se restringe a problemas com respostas contínuas que tenham distribuições normais ou que possam se aproximar de distribuições normais após alguma transformação. Supõe-se, também, que as variâncias das respostas possam ser representadas por modelos lineares após uma transformação logarítmica. Observa-se que estas duas suposições são usuais na literatura sobre o assunto.

REVISÃO DA LITERATURA

Este capítulo faz uma breve descrição sobre a metodologia de superfície de resposta e explora a questão da modelagem da variância da resposta. Considera o problema de modelar simultaneamente médias e variâncias de múltiplas características de qualidade e conclui com a descrição de alguns métodos de otimização com múltiplos critérios.

2.1 - METODOLOGIA DE SUPERFÍCIE DE RESPOSTA

A metodologia de superfície de resposta (RSM) pode ser entendida como uma combinação de técnicas de planejamento de experimentos, análise de regressão e métodos de otimização. Tem larga aplicação nas pesquisas industriais, particularmente em situações onde um grande número de variáveis de um sistema influencia alguma característica fundamental deste sistema.

Box et al. (1978) apresentam a RSM como uma metodologia que consiste num grupo de técnicas usadas em estudos empíricos. Estas técnicas relacionam uma ou mais respostas, tais como produção, tonalidade e viscosidade com variáveis de entrada, tais como tempo, temperatura, pressão e concentração. A RSM tem sido usada para responder questões do tipo:

- a) Como uma resposta particular é afetada por um conjunto de variáveis de entrada sobre uma região de interesse?
- b) Quais conjuntos de entradas geram um produto satisfazendo simultaneamente várias especificações desejadas?
- c) Quais os valores de entradas que produzem o valor máximo para uma resposta específica e como se comporta a resposta em torno deste máximo?

Dado um processo ou sistema, com várias variáveis (ou fatores) de entrada, $x = (x_1, x_2, \dots, x_k)'$ e uma variável de saída (ou resposta) y , a RSM normalmente consiste em:

- a) planejar um experimento que permita estimar uma equação de regressão (superfície de resposta) do tipo

$$E(y / x) = f(x, \beta), \quad x \in D \subset \mathcal{R}^k \quad (2.1)$$

onde: $E(y / x)$ representa o valor esperado de y , condicionado a pontos $x \in D$;

f é uma função (usualmente um polinômio de grau não superior a 2);

x é o vetor de variáveis de entrada;

β é um vetor de parâmetros a ser estimado a partir dos dados do experimento e

$D \subset \mathcal{R}^k$ é a região especificada para a investigação.

- b) com os dados do experimento, estimar o vetor de parâmetros β , avaliando a significância estatística de seus componentes e
- c) estudar o comportamento da função f (por exemplo, se é desejado o maior valor possível para a resposta y , um dos interesses é obter $x_0 \in D$ que maximiza f , caso ele exista).

A revista *Technometrics* apresenta dois trabalhos bastante amplos de revisão da literatura sobre RSM: Hill e Hunter (1966) e Myers et al. (1989). O segundo abrange a questão do planejamento e da análise de experimentos (incluindo situações de múltiplas respostas), áreas de aplicação e orientações para futuras pesquisas.

2.1.1 - Planejamento e análise de experimentos

Conforme descrito por Montgomery (1997, p.1), experimentos são realizados por investigadores das mais variadas áreas, usualmente para descobrir alguma coisa sobre um processo ou sistema particular. Literalmente, um experimento é um teste. Um experimento planejado é um teste, ou uma série de testes, nos quais são feitas alterações planejadas nas variáveis de entrada do processo ou sistema, permitindo observar e identificar os efeitos das mudanças na variável de saída (resposta).

As idéias básicas da teoria do planejamento de experimento e a sua relação com técnicas estatísticas, tais como a análise de variância, foram desenvolvidas por R. A. Fisher nas décadas de 20 e 30. A maioria dos experimentos, nesta época, voltavam-se para a agricultura e biologia. Realizavam-se experimentos para verificar, por exemplo, o efeito de diferentes níveis e tipos de fertilizantes (variáveis controláveis de entrada) sobre a produção de uma determinada cultura (variável de saída ou resposta).

Alguns livros mostram a evolução nesta área. Fisher (1951) descreve os princípios básicos da experimentação, Cochran e Cox (1957) apresentam a abordagem clássica de forma bastante didática e com vários exemplos aplicados à agricultura e à biologia e Box et al. (1978) enfatizam as aplicações na área industrial. Taguchi (1987) é uma tradução para o inglês do livro originalmente lançado no Japão em 1976. Neste livro, Taguchi apresenta algumas técnicas de planejamento de experimento com novo enfoque e tendo como público alvo os engenheiros. Myers e Montgomery (1995) e Montgomery (1997) apresentam o planejamento e análise de experimentos como uma forma de promover a melhoria da qualidade. Seguindo os conceitos de engenharia de Taguchi, estes dois últimos livros procuram enfatizar o efeito dos fatores de entrada não apenas sobre o valor esperado da resposta, mas também sobre a sua variância.

Coleman e Montgomery (1993) enfatizam as fases iniciais de um planejamento de experimento. Ressaltam a necessidade do claro entendimento do problema por parte dos engenheiros e estatísticos, da escolha adequada dos fatores (variáveis de entrada), da identificação da amplitude dos níveis dos fatores que serão ensaiados e da seleção das respostas (medidas de desempenho do produto). Caten e Ribeiro (1996) também abordam este tema de forma bastante didática. Observam que o estudo experimental deve ser voltado para o cliente e, sendo assim, a escolha das respostas deve ser feita em termos de características de qualidade levantadas junto aos clientes.

Myers e Montgomery (1995, p. 10) observam que muitos problemas de RSM são sequenciais por natureza. Geralmente, numa fase inicial, a quantidade de possíveis fatores é muito grande, necessitando de um experimento preliminar (*screening experiment*) para eliminar fatores não significativos. A região convexa

compreendida entre os níveis ensaiados dos fatores é chamada de região experimental. Numa fase seguinte, procura-se encontrar a região onde se localiza a melhor combinação de níveis dos fatores, podendo-se levantar uma nova região experimental para a investigação. E, finalmente, realiza-se um experimento com mais ensaios para se estimar a superfície de resposta sobre a região experimental.

Para o entendimento do restante desta seção, fazem-se necessários alguns conceitos. Denomina-se *efeito principal* de um fator à diferença provocada na resposta, quando este fator muda do nível mais baixo para o nível mais alto.

Diz-se que existe *interação* entre dois fatores quando a diferença na resposta entre os níveis de um fator não é a mesma para todos os níveis do outro fator. No caso, é uma interação de segunda ordem, porque envolve dois fatores. De forma análoga, pode-se definir interações de ordem mais alta, mas normalmente os efeitos de interações de ordem superior a dois são pequenos e de difícil interpretação prática. Observa-se que quando não existe interação entre os fatores, os seus efeitos são *aditivos*.

O termo *confundimento* entre dois ou mais fatores será usado para as situações onde o projeto experimental não permite avaliar o efeito de cada fator isoladamente. O confundimento pode ser parcial ou total. Por exemplo, se todos os ensaios são realizados ou com temperatura e pressão baixas, ou com temperatura e pressão altas, haverá um confundimento total entre os fatores pressão e temperatura.

Na fase inicial da investigação, onde o número de fatores pode ser bastante grande, torna-se inviável a realização de um planejamento fatorial completo, o qual consiste em combinar todos os níveis dos fatores. Por exemplo, se existem oito fatores e cada um deles é observado em dois níveis, precisa-se de $2^8 = 256$ ensaios. Em três níveis, $3^8 = 6.561$ ensaios. Isto sem contar com as repetições de ensaios sob as mesmas condições, ou *replicações*, que são úteis para avaliar a variância do processo. Quando o número de fatores é bastante grande, a RSM orienta iniciar a pesquisa com os fatores ensaiados em apenas dois níveis e com planejamentos fatoriais fracionados. Estes planejamentos são denotados por 2^{k-r} , onde k é o número de fatores e r é o número de fracionamentos. Nos planejamentos fracionados os efeitos das interações de ordem alta são confundidos entre si.

Sempre que possível, procura-se trabalhar com projetos de experimentos que não produzam confundimento entre os efeitos de cada fator isoladamente (efeitos principais) e de interações de segunda ordem. Por exemplo, num experimento com oito fatores, ensaiados com dois níveis cada um, precisa-se de $2^{8-2} = 64$ ensaios para garantir estimativas isoladas para todos os efeitos principais e interações de segunda ordem (ver Montgomery, 1997, p. 689). Um projeto de experimento fatorial fracionado pode ser gerado a partir de uma relação inicial de confundimento. Os programas de estatística (SAS, STATISTICA, STATGRAPHICS, etc.) normalmente fornecem o projeto de experimento resultante e todos os confundimentos conseqüentes (ver, também, Montgomery, 1997, p. 372 e 683).

O conhecimento de engenharia do produto ou processo permite descartar certas interações, o que pode reduzir drasticamente o número de ensaios. Uma linha de trabalho, conhecida como métodos de Taguchi, leva ao extremo o conhecimento prévio do produto ou processo para planejar um experimento. Nesta abordagem, é fundamental especificar, *a priori*, todas as interações que possam influenciar a resposta. Em situações onde não existem interações, pode-se usar até sete fatores com apenas oito ensaios. Ross (1991, p. 308) exemplifica um estudo com seis fatores e uma interação num projeto experimental com apenas oito ensaios.

Com os dados resultantes do experimento, a identificação dos efeitos significativos pode ser feita através de técnicas estatísticas formais (testes t ou F da análise de variância), ou por métodos visuais. Para realizar os testes estatísticos da análise de variância, necessita-se estimar a variância do erro. Ela pode ser estimada quando o experimento tem replicações (ver Box et al., 1978, seção 10.6 ou Montgomery, 1997, seção 6.4), quando houver ensaios adicionais no ponto central, isto é, na combinação do ponto médio de todos os fatores (ver Montgomery, 1997, p. 580) ou, ainda, agregando os efeitos de menor magnitude ao termo de erro (ver Nair e Pregibon, 1986 e Chao e Hwang, 1997). Este último critério, apesar de ser bastante usado, pode ser viciado (ver Box, 1988).

Muitos autores preferem métodos menos formais para separar os efeitos mais relevantes, tais como o gráfico de Pareto e o gráfico normal dos efeitos (*normal plot of effects*). Este último relaciona a magnitude de cada efeito (principal ou de

interação) com a distribuição normal acumulada. Os efeitos não significativos tendem a ficar em torno de uma reta que passa pela origem, enquanto os significativos ficam nas extremidades e fora da linha da reta (ver exemplos nos capítulos 3 e 5). O gráfico normal dos efeitos pressupõe que a maioria dos fatores produzem efeitos nulos (princípio da *esparsidade*). Além disso, pressupõe que a resposta tenha distribuição normal com mesma variância, mas é bastante robusto a cerca desta última suposição (Asscher e Feigin, 1993 apud Engel e Huele, 1996). Alternativamente, pode-se usar os valores absolutos dos efeitos, produzindo o *half-normal plot of effects*. Este gráfico permite visualizar melhor os efeitos significativos.

Vários artigos procuram tratar mais formalmente as idéias do gráfico normal dos efeitos, propondo testes estatísticos para a verificação dos efeitos significativos em experimentos sem replicações. Box e Meyer (1986b) utilizam-se da abordagem bayesiana. Alguns parâmetros da distribuição *a priori*, como a proporção de efeitos significativos, são sugeridos com base num levantamento de dezenas de análises de experimentos. Lenth (1989) sugere um procedimento mais simples, onde se constróem intervalos de confiança simultâneos para os vários efeitos, baseando-se num pseudo erro padrão. Esta estatística é calculada através de uma abordagem robusta, a qual só inclui os efeitos não discrepantes. Olguín e Fearn (1997) propõem uma nova estatística para o teste, com distribuição gerada por simulação Monte Carlo. A distribuição simulada, para os casos mais comuns, é apresentada no artigo.

Depois de identificados os fatores significativos, o experimento pode ser aumentado com ensaios adicionais no centro da região experimental. Isto permite avaliar se a superfície de resposta linear é adequada ou se existe alguma curvatura (ver Myers e Montgomery, 1995, p. 111). O teste de curvatura serve para verificar se o nível ótimo de operação encontra-se na região de investigação. Se o teste não detectar curvatura, pode-se realizar novos experimentos para se encontrar a região onde se localiza o ponto ótimo.

O chamado método da máxima inclinação ascendente permite mover o experimento rapidamente para a vizinhança geral do ótimo, utilizando-se um procedimento experimental simples, rápido e eficiente (Riboldi e Nascimento, 1994, p. 9).

Determinados os principais fatores que influenciam a resposta e a região de investigação, busca-se normalmente um projeto de experimento que permita estimar uma superfície de resposta formada por um polinômio de segundo grau. Supõe-se, aqui, que os fatores são contínuos, tais como temperatura e pressão. Um planejamento do tipo central composto (CCD) geralmente é o mais adequado para o ajuste de uma função de segundo grau. Ele é formado pelo planejamento fatorial 2^k , acrescido de ensaios nos pontos central e axiais, como ilustra a figura 2.1 para $k = 2$.

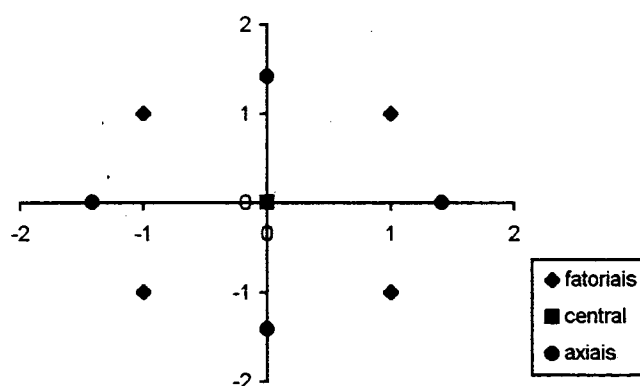


Fig. 2.1 - Ilustração de um CCD com dois fatores e níveis codificados na forma padrão.

A codificação dos níveis dos fatores é discutida em Myers e Montgomery (1995, p. 22). Riboldi e Nascimento (1994, p. 46 e 47) apresentam os valores de α , relativos aos pontos axiais $(0, \dots, 0, \alpha, 0, \dots, 0)$ e o número adequado de ensaios no ponto central, que levam a propriedades desejáveis do projeto de experimento.

É muito comum, na área industrial, utilizar-se de projetos de experimentos com alguns fatores ensaiados em dois níveis e outros em três níveis, tal como o arranjo ortogonal L_{18} (ver Taguchi, 1987 ou Ross, 1991). Sobre os resultados de um experimento deste tipo, pode-se construir uma superfície de resposta que inclui alguns termos quadráticos e algumas interações.

2.1.2 - Construção de modelos

Com os dados do experimento, busca-se estimar uma superfície de resposta, que relaciona uma resposta do processo com os fatores de entrada. Isto é feito através da *análise de regressão*. Conforme descreve Stigler (1986), o termo *regressão* surgiu com os trabalhos de Galton, no final do século passado. Estes trabalhos procuravam, com observações empíricas, explicar certas características de

um indivíduo a partir das características de seus pais, donde o modelo estatístico utilizado ficou conhecido com o nome de *regressão*.

Na RSM, procura-se estabelecer uma superfície de resposta do tipo

$$E(y/x) = f(x, \beta), \quad x \in D \subset \mathbb{R}^k \quad (2.2)$$

onde: $E(y/x)$ representa o valor esperado da resposta y , condicionado a pontos $x \in D$;

f é uma função matemática;

x é o vetor de variáveis de entrada, $x = (x_1, x_2, \dots, x_k)'$;

β é um vetor de parâmetros a ser estimado com os dados do experimento e

$D \subset \mathbb{R}^k$ é a região experimental.

Uma superfície de primeira ordem (ou linear) tem a seguinte forma:

$$f(x, \beta) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (2.3)$$

E uma superfície de segunda ordem (ou quadrática) pode ser escrita como

$$f(x, \beta) = \underbrace{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}_{\text{termos lineares}} + \underbrace{\beta_{11} x_1^2 + \dots + \beta_{kk} x_k^2}_{\text{termos quadráticos puros}} + \underbrace{\beta_{12} x_1 x_2 + \dots + \beta_{(k-1)k} x_{k-1} x_k}_{\text{termos de interação}} \quad (2.4)$$

Quando se modela a resposta do i -ésimo ensaio, y_i ($i = 1, 2, \dots, n$), deve-se considerar os ruídos gerados pela infinidade de fatores não incluídos no modelo e que agem de forma não sistemática sobre a resposta. Os ruídos produzirão um termo adicional na equação (2.2), o qual é chamado de *erro aleatório* ou, simplesmente, de *erro* e será denotado por e_i ($i = 1, 2, \dots, n$). Admite-se que os erros e_i ($i = 1, 2, \dots, n$) satisfaçam as seguintes propriedades:

- sejam aditivos à função $f(x_i, \beta)$;
- sejam independentes;
- tenham valor esperado nulo;
- tenham variância constante e
- preferencialmente, tenham distribuição normal de probabilidade.

Uma superfície de resposta linear pode ser escrita em termos das observações y_i ($i = 1, 2, \dots, n$), como

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i \quad (2.5)$$

E uma superfície de resposta quadrática, como

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \beta_{11} x_{i1}^2 + \dots + \beta_{kk} x_{ik}^2 + \beta_{12} x_{i1} x_{i2} + \dots + \beta_{(k-1)k} x_{i(k-1)} x_{ik} + \varepsilon_i \quad (2.6)$$

Ambas as superfícies de resposta podem ser escritas em termos de um modelo linear geral, tal como

$$y = X \beta + \varepsilon \quad (2.7)$$

onde: y é um vetor $n \times 1$, com as observações da variável resposta;

X é uma matriz $n \times p$, com as variáveis de entrada (matriz de planejamento);

β é um vetor $p \times 1$, com os parâmetros do modelo;

ε é um vetor $n \times 1$, com os erros aleatórios.

No modelo quadrático completo, por exemplo, a matriz X é formada por uma coluna unitária, referente ao termo constante, k colunas compostas pelos valores de x_1, x_2, \dots, x_k , k colunas com os termos quadráticos puros, $x_1^2, x_2^2, \dots, x_k^2$ e $k(k-1)/2$ colunas formadas pelas interações $x_1 x_2, \dots, x_{k-1} x_k$. Logo, sua dimensão será $n \times p$, com $p = 1 + 2k + [k(k-1)/2]$. Em geral, usam-se como x_1, x_2, \dots, x_k valores codificados de tal forma que ficam centrados em zero e com mesma dispersão. Os valores codificados devem ter relações lineares com os valores originais dos fatores (ver, por exemplo, Khuri e Cornell, 1996, p. 47).

A estimação dos parâmetros

Dada a forma do modelo, necessita-se estimar o vetor de parâmetros β . O método mais conhecido é o dos *mínimos quadrados* (MQO), que busca o vetor $\hat{\beta}$ que minimiza a soma quadrática dos erros, ou seja, $\hat{\beta}$ é o resultado da seguinte operação matricial:

$$\min_{\beta} (\varepsilon' \varepsilon) = \min_{\beta} \{(y - X\beta)'(y - X\beta)\}$$

cujas solução é

$$\hat{\beta} = (X'X)^{-1} X'y \quad (2.8)$$

O estimador $\hat{\beta}$, além de ser *não-viciado*¹, goza da seguinte propriedade, conhecida como Teorema de Gauss - Markov (ver, por exemplo, Judge et al., 1988,

¹ Um estimador é dito *não-viciado* se o seu valor esperado coincide com o parâmetro que se deseja estimar.

p. 203): *dentre todos os estimadores lineares não-viciados, o estimador de mínimos quadrados é aquele que tem a menor variância, isto é, o mais eficiente.*

Outra característica interessante é que, se for admitido que o termo de erro segue uma distribuição normal, então o estimador de (2.8) é também o estimador de *máxima verossimilhança*. Pelo método da máxima verossimilhança, o estimador $\hat{\beta}$ é aquele que maximiza a probabilidade (verossimilhança) da distribuição subjacente gerar uma amostra igual aquela que foi observada. Maiores detalhes sobre este método de estimação pode ser visto em Cordeiro (1992).

Obtidas as estimativas dos parâmetros, torna-se conveniente avaliar se a forma funcional do modelo, previamente estabelecida, é adequada. Isto pode ser feito por um gráfico que relaciona as estimativas dos erros com os correspondentes valores preditos, chamado *gráfico dos resíduos*.² O vetor de estimativas dos erros (ou *resíduos*) e o vetor dos valores preditos são calculados, respectivamente, por

$$\hat{\varepsilon} = y - X\hat{\beta} \quad \text{e} \quad \hat{y} = X\hat{\beta} \quad (2.9)$$

Se o modelo foi especificado adequadamente, os elementos de $\hat{\varepsilon}$ devem apresentar-se aleatoriamente em torno de zero, independentemente da magnitude dos valores preditos. Além disso, é desejável que os resíduos sigam a forma de uma distribuição normal. Detalhes deste gráfico e formas alternativas são encontrados em livros de análise de regressão, tais como Chatterjee e Price (1977), Drapper e Smith (1981) e Carroll e Ruppert (1988).

A métrica da resposta

Um modelo pode não se ajustar bem aos dados devido à métrica usada para a resposta. Por exemplo, ao avaliar uma chapa de aço pelo número de micro cavidades por m^2 , está se supondo que duas chapas com, respectivamente, 20 e 22 micro cavidades por m^2 são tão diferentes quanto outras duas, com 2 e 4 micro cavidades por m^2 , pois em ambos os casos a diferença é de duas unidades. Numa situação como esta, talvez seja mais adequado se trabalhar com uma transformação da resposta, a

² Esta avaliação também pode ser feita por testes estatísticos, como será discutido posteriormente.

qual aumente a distância entre os valores pequenos e reduza as distâncias entre os valores grandes, tais como \sqrt{y} ou $\log(y)$, $y > 0$.

Uma transformação da resposta pode ser efetuada para se obter:

- a) adequação de uma função simples para a superfície de resposta;
- b) variância constante (homocedasticidade) na região experimental e
- c) distribuição normal para os erros.

Chatterjee e Price (1977, p. 29) discutem transformações para se obter um modelo linear para a resposta. Bartlett e Kendall (1946) e Chatterjee e Price (1977, p. 38) apresentam transformações para estabilizar a variância. Johnson e Wichern (1992, p. 164) sugerem algumas transformações para se aproximar da distribuição normal. De modo mais amplo, a RSM usa muito a chamada *transformação de Box-Cox*, que procura atingir simultaneamente os itens (a), (b) e (c).

A transformação de Box-Cox utiliza a família de funções

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda \bar{y}^{\lambda-1}}, & \text{para } \lambda \neq 0 \\ \bar{y} \log(y), & \text{para } \lambda = 0 \end{cases} \quad (2.10)$$

onde \bar{y} é a média geométrica dos dados e “log” representa o logaritmo neperiano. O valor de λ é escolhido de forma a minimizar o erro quadrático médio (ver, por exemplo, Box, 1978, p. 219).

Myers e Montgomery (1995, p. 260) alegam que o sucesso da técnica de Box - Cox depende da amplitude das observações, sugerindo realizá-la quando a razão entre o maior e o menor valor for superior a três. Supõe-se aqui que os valores de y são sempre positivos. Judge et al. (1988, p. 560) mostram, através de um estudo de simulação, a fragilidade da técnica, numa situação em que a amplitude das observações é pequena. A utilização da transformação de Box-Cox também pode ficar prejudicada na presença de valores discrepantes. Diversos autores, dentre eles, Carroll e Ruppert (1985), Atkinson (1986) e Kim et al. (1996), chamam a atenção para este fato e propõem algumas alterações da técnica.

Chatterjee e Price (1977, p. 37) observam que as propriedades dos estimadores num modelo linearizado não são preservadas quando se efetua a

transformação inversa. Por exemplo, se o modelo $\log(y) = \alpha + \beta x$ é adequado para representar a variação sistemática da resposta y , os estimadores de mínimos quadrados, $\hat{\alpha}$ e $\hat{\beta}$, satisfazem as propriedades de serem *não-viciados* e do Teorema de Gauss - Markov. Contudo, ao colocar o modelo estimado em termos da resposta, $\hat{y} = \exp\{\hat{\alpha} + \hat{\beta}x\}$, o fator constante, $\hat{\gamma} = \exp\{\hat{\alpha}\}$, é um estimador *viciado* do parâmetro $\gamma = \exp\{\alpha\}$.

Testes no modelo de regressão

Ao estabelecer um modelo linear para a resposta, ou para uma transformação da resposta, normalmente não se sabe, *a priori*, quais os termos que devem fazer parte do modelo. Por isto, é comum realizar testes estatísticos sobre grupos de coeficientes ou sobre cada coeficiente individualmente. Os testes descritos a seguir pressupõem que os termos de erro seguem uma distribuição normal de probabilidade.

O teste mais geral é o teste F da análise da variância, que verifica a significância do modelo como um todo. A hipótese nula atesta que todos os coeficientes são nulos e, portanto, a sua rejeição é fundamental para que se continue trabalhando em termos da família de modelos estabelecida e dos fatores de entrada selecionados. A estatística deste teste é dada por

$$F_0 = \frac{SS_R/(p-1)}{SS_E/(n-p)} \quad (2.11)$$

onde: p é o número de parâmetros do modelo;

n é o número de observações;

$SS_E = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - \hat{\beta}'X'y$, isto é, a soma de quadrados devida ao erro e

$$SS_R = \hat{\beta}'X'y - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}, \quad \text{ou seja, a soma de quadrados devida à regressão;}$$

Sob a hipótese nula, F_0 tem distribuição F com $(p-1)$ graus de liberdade no numerador e $(n-p)$ graus de liberdade no denominador. Logo, o teste rejeita a

hipótese nula se F_0 exceder $F_{\alpha, p-1, n-p}$, onde α é o nível de significância adotado (por exemplo, $\alpha = 0,05$) e $F_{\alpha, p-1, n-p}$ é a correspondente abscissa da distribuição F.

Nas situações onde o experimento tem replicações, ou seja, mais de um ensaio nas mesmas condições experimentais, torna-se possível realizar um teste estatístico para a falta de ajustamento. O teste baseia-se na decomposição da soma de quadrado dos erros (SS_E) em erro puro (SS_{PE}) e falta de ajustamento (SS_{LOF}). Sejam:

M o número de diferentes condições experimentais;

m o número de replicações;

$n = Mm$;

p o número de parâmetros do modelo;

\hat{y}_i ($i = 1, \dots, m$) o valor predito pela regressão em cada ponto experimental;

\bar{y}_i ($i = 1, \dots, m$) a média aritmética da resposta em cada ponto experimental;

$$SS_{PE} = \sum_{i=1}^M \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2$$

$$SS_{LOF} = m \sum_{i=1}^M (\bar{y}_i - \hat{y}_i)^2$$

A estatística do teste é dada por

$$F_0 = \frac{SS_{LOF}/(M - p)}{SS_{PE}/(n - M)} \quad (2.12)$$

e o teste rejeita o modelo por falta de ajustamento quando $F_0 > F_{\alpha, M-p, n-M}$.

A significância de grupos de coeficientes também pode ser avaliada por um teste F. Este teste é importante para verificar, por exemplo, se existe necessidade de se considerar uma superfície de resposta quadrática. Nesta situação, considera-se no modelo os termos lineares e testa-se se os coeficientes dos termos quadráticos podem ser admitidos como nulos. Para testar a hipótese nula de que o vetor de parâmetros β_1 , com r elementos, pode ser nulo, dado que o vetor β_2 já esteja contemplado no modelo, aplica-se a estatística

$$F_0 = \frac{SS_R(\beta_1|\beta_2)/r}{SS_E/(n - p)} \quad (2.13)$$

onde $SS_R(\beta_1|\beta_2)$ é o incremento na soma de quadrados devido à inclusão de β_1 . O teste rejeita a hipótese nula quando F_0 exceder $F_{\alpha, r, n-p}$.

Uma hipótese mais genérica, conhecida como hipótese linear geral, é apresentada em Judge et al. (1988, p. 252). Em termos matriciais, as hipóteses são $H_0: R\beta = r$ vs. $H_1: R\beta \neq r$, contemplando J ($1 \leq J < p$) relações entre os parâmetros do modelo, onde:

β é o vetor $p \times 1$ de parâmetros do modelo;

R é uma matriz $J \times p$ que especifica relações entre os parâmetros e

r é um vetor $J \times 1$ de valores numéricos, completando as relações entre os parâmetros.

Por exemplo, para testar $H_0: \beta_1 = \beta_2$, num modelo linear com vetor de parâmetros $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_{(p-1)})'$, fazem-se $R = (0, 1, -1, 0, \dots, 0)$ e $r = 0$, donde tem-se a relação $R\beta = r \Rightarrow \beta_1 - \beta_2 = 0 \Rightarrow \beta_1 = \beta_2$.

A hipótese linear geral pode ser testada pela estatística

$$F_0 = \frac{(R\hat{\beta} - r)' [R(X'X)^{-1}R']^{-1} (R\hat{\beta} - r)}{J\hat{\sigma}^2} \quad (2.15)$$

Sob a hipótese nula, F_0 tem distribuição F com J e $(n - p)$ graus de liberdade. Logo, o teste rejeita a hipótese nula se F_0 exceder $F_{\alpha, J, n-p}$. Em termos computacionais, este tipo de hipótese pode ser especificada no SAS (*Statistical Analysis System*), proc GLM, por exemplo.

Um teste mais específico avalia a significância de cada termo do modelo. A hipótese $H_0: \beta_j = 0$ versus $H_1: \beta_j \neq 0$, para algum $j = 1, \dots, p$, pode ser testada pela estatística

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 c_{jj}}} \quad (2.14)$$

onde: c_{jj} é o elemento da diagonal de $(X'X)^{-1}$ correspondente ao elemento β_j e

$$\hat{\sigma}^2 = SS_E / (n - p).$$

A hipótese nula é rejeitada quando t_0 exceder $t_{\alpha, n-p}$, onde α é o nível de significância adotado e $t_{\alpha, n-p}$ é a correspondente abscissa da distribuição t de *Student* com $(n - p)$ graus de liberdade. Na medida do possível, deve-se dar preferência a testes mais gerais do tipo F , ao invés de se efetuar vários testes do tipo t .

Os testes estatísticos são procedimentos bastante formais e, por isto, são bastante influenciados quando as suposições do modelo são violadas. A presença de valores discrepantes, por exemplo, pode alterar totalmente os resultados dos testes. Neste sentido, abordagens gráficas podem ser usadas como alternativas ou complementos dos testes. Além dos gráficos citados na seção 2.1.1, Cook e Weisberg (1989) e Cook (1994) discutem a utilização de gráficos para verificar a influência de novas variáveis num modelo de regressão, incluindo recursos de animação, onde pode-se acompanhar suavemente a tendência nos resíduos quando novas variáveis são incluídas no modelo.

Dado o ajuste de um modelo, é comum avaliar a sua capacidade preditiva. Tradicionalmente isto é feito pelo coeficiente de determinação R^2 , que fornece a proporção da variabilidade da resposta que pode ser explicada pelos fatores incluídos no modelo, admitindo correta sua forma. Este coeficiente pode ser definido da seguinte maneira:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \quad (2.16)$$

$$\text{onde } SS_T = y'y - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n}.$$

Para se levar em conta o custo de incluir termos adicionais, pode-se descontar os graus de liberdade perdidos com cada termo do modelo, resultando no coeficiente de determinação ajustado, dado por³

$$R_{aj}^2 = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} \quad (2.17)$$

Seleção das variáveis independentes

Quando se está trabalhando com dados gerados por um experimento central composto de segunda ordem, tem-se observações para estimar todos os coeficientes de uma superfície de resposta quadrática. Muitos autores, dentre eles Vining e Myers (1990) e Ames et al. (1997), sugerem usar o modelo estimado completo, pois na

³ Observa-se que as relações (2.11), (2.16) e (2.17) estão interligadas, pois $F_0 = \frac{R^2/(p-1)}{(1-R^2)/(n-p)}$ e

$$R_{aj}^2 = 1 - \frac{n-1}{n-p}(1-R^2).$$

RSM o modelo deve ter boa qualidade preditiva e a exclusão de termos normalmente piora a qualidade de predição e pode provocar vícios nas estimativas. Por outro lado, muitas vezes a preocupação maior é encontrar modelos simples (parcimoniosos), onde a inclusão de termos não significativos seria inadequada (ver Ribeiro et al., 1997).

Uma abordagem usada neste trabalho (ver capítulos 3 e 5) consiste em aplicar sucessivos testes F. Aplica-se o teste F para testar a significância do modelo quadrático completo usando a estatística (2.11). Se significativo, testa-se seqüencialmente os seguintes subgrupos de coeficientes: (1) dos termos lineares, (2) dos termos lineares e quadráticos puros e (3) dos termos lineares e interações de segunda ordem. Nos dois últimos casos, utiliza-se a estatística (2.13).

Uma outra forma é escolher os termos mais significativos é através de um método de seleção automática de variáveis, tal como o método baseado na estatística C_p de Marlow (ver Guedes, 1996 e Myers e Montgomery, 1995, p. 640).

Um tratamento especial deve ser dado aos planejamentos experimentais onde existem confundimentos parciais entre os efeitos, como o de Plackett-Burman (ver Myers e Montgomery, 1995, p. 169) ou mesmo nos arranjos ortogonais L_{18} e L_{36} sugeridos por Taguchi (1987). No segundo caso, os efeitos quadráticos e de interação de segunda ordem são parcialmente confundidos. Os dados gerados por estes experimentos não permitem estimar todos os parâmetros de uma superfície quadrática, necessitando de técnicas alternativas para escolher quais parâmetros devem ser incluídos no modelo. Hamada e Wu (1992), Chao e Hwang (1997) e Chipman et al. (1997) apresentam propostas para este problema.

Descreve-se, a seguir, a proposta de Hamada e Wu (1992), que baseia-se nos princípios da *esparsidade* e da *hereditariedade*. Estes princípios foram descritos em Box e Meyer (1986b). O princípio da *esparsidade* é uma extensão do princípio tradicional de Pareto: “os poucos vitais e os muitos triviais”, para o problema de identificar os principais efeitos dentre um grande número de possibilidades de combinações de níveis dos fatores. Já o princípio da *hereditariedade* considera que as interações só fazem sentido físico se os efeitos principais dos fatores envolvidos na interação são significativos.

O método proposto por Hamada e Wu (1992) consiste no seguinte algoritmo:

Passo 1 - Consideram-se, inicialmente, todos os efeitos principais e selecionam-se os efeitos significativos, usando testes estatísticos ou métodos gráficos.

Passo 2 - Usando-se o efeito da hereditariedade, considera-se (a) os efeitos identificados no passo 1, (b) as interações de segunda ordem que tenham pelo menos um componente dentre os efeitos principais significativos e (c) interações sugeridas pelo experimentador. Selecionam-se os efeitos significativos deste grupo através de uma regressão em que as variáveis entram no modelo, uma de cada vez (*forward regression*).

Passo 3 - Utiliza-se novamente *forward regression* com os efeitos significativos do passo 2 e todos os efeitos principais.⁴

Os passos 2 e 3 devem ser repetidos até que não haja mais alterações no modelo. Este algoritmo foi usado em estudos de caso por Hamada e Wu (1992) e por Mesenbrink et al. (1994).

2.1.3 - Análise da superfície de resposta e otimização

Com o modelo de regressão ajustado aos dados experimentais, pode-se obter várias informações sobre o comportamento da resposta do sistema em função dos fatores controláveis do processo. Por exemplo, através do sinal e da magnitude relativa dos coeficientes, pode-se prever a variação da resposta a partir de mudanças planejadas nos fatores. Mas o maior interesse, na maioria das vezes, é obter a otimização do processo em termos dos fatores controláveis, assim como o conhecimento da sensibilidade da resposta em torno da posição ótima.

Existe um conjunto bastante grande de métodos para o problema de se encontrar $x_0 \in D$ que maximiza (ou minimiza) a resposta, admitindo-se a sua existência (ver, por exemplo, Bazaraa, 1993). Em especial, se os fatores x_1, x_2, \dots, x_k são contínuos e podem variar livremente em D e, também, se as equações de regressão são formadas por funções deriváveis, o método baseado no gradiente permite realizar uma análise bastante informativa. A seguir, descreve-se um pequeno

⁴ O método de Hamada e Wu não cita os efeitos quadráticos, mas eles podem ser incluídos no passo 1.

resumo deste método. Myers e Montgomery (1995, cap. 6) e Riboldi e Nascimento (1994, cap. 3) apresentam uma exposição mais detalhada.

Seja o modelo estimado:

$$\hat{y} = X\hat{\beta} \quad (2.18)$$

que também pode ser apresentado em termos da média, $\hat{\mu}_y = X\hat{\beta}$, considerando que a parte preditiva de y , segundo o modelo (2.1), é o seu valor esperado.

Assumindo que (2.18) seja um modelo quadrático completo, pode-se escrevê-lo da seguinte maneira:

$$\hat{y} = b_0 + x'b + x'Bx \quad (2.19)$$

onde: $x = (x_1, x_2, \dots, x_k)'$ é um vetor $k \times 1$, formado pelos fatores de entrada;

b_0 é o termo constante;

$b = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)'$ é um vetor $k \times 1$, formado pelos termos lineares e

$$B = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \dots & \hat{\beta}_{1k}/2 \\ & \hat{\beta}_{22} & \dots & \hat{\beta}_{2k}/2 \\ \text{sim.} & & \dots & \\ & & & \hat{\beta}_{kk} \end{bmatrix}$$

é uma matriz $k \times k$, simétrica, formada pelos coeficientes dos termos quadráticos puros (diagonal principal) e pelos coeficientes das interações.

Fazendo-se $\frac{\partial \hat{y}}{\partial x} = 0$, encontra-se, como solução, o seguinte ponto estacionário:

$$x_s = -\frac{1}{2}B^{-1}b \quad (2.20)$$

que pode ser um ponto de máximo, um ponto de mínimo ou um ponto de sela. A sua natureza é determinada pelos autovalores $\lambda_1, \dots, \lambda_k$ da matriz B , da seguinte forma:

- a) se $\lambda_i < 0$, para todo $i = 1, \dots, k$, então x_s é um ponto de máximo;
- b) se $\lambda_i > 0$, para todo $i = 1, \dots, k$, então x_s é um ponto de mínimo e
- c) se existe $\lambda_i < 0$ e $\lambda_j > 0$ $i, j = 1, \dots, k$, então x_s é um ponto de sela.

A superfície de resposta pode ser colocada na forma canônica, como segue:

$$\hat{y} = \hat{y}_s + \lambda_1 w_1^2 + \lambda_2 w_2^2 + \dots + \lambda_k w_k^2 \quad (2.21)$$

onde: \hat{y}_s é o valor de \hat{y} no ponto estacionário x_s e

$w = (w_1, w_2, \dots, w_k)' = P'(x - x_s)$, sendo P uma matriz $k \times k$, formada pelos autovetores normalizados da matriz B .

A forma canônica permite conhecer a natureza do sistema em torno do ponto estacionário. Por exemplo, suponha que $k = 2$ e deseja-se encontrar a combinação dos dois fatores que maximiza a resposta. Na análise, verifica-se que os dois autovalores de B são negativos, com λ_2 muito próximo de zero. Então, tem-se que o ponto estacionário x_s é um ponto de máximo e, além disso, devido λ_2 ser próximo de zero, pode-se obter uma reta L dentro do plano de investigação, tal que a resposta mantém-se próxima do máximo. Desta forma, conclui-se que o processo pode ser operado em qualquer combinação de níveis de x_1 e x_2 que pertença à reta L . Dentre estes níveis, pode-se escolher aquele que tem o menor custo de operação.

Os programas computacionais que realizam a análise de superfícies de resposta (SAS, STATGRAPHICS, STATISTICA, etc.) geralmente são munidos do chamado gráfico de contornos. Em sua forma mais simples, o gráfico de contornos apresenta as curvas de níveis da superfície de resposta em função de cada par dos fatores de entrada, sendo os demais mantidos constantes. Desta forma, torna-se possível visualizar a região em torno do nível ótimo de operação, conhecendo-se quais fatores precisam ser cuidadosamente fixados em seus níveis ótimos e quais podem variar sem muito prejuízo para o sistema.

2.2 - CONSTRUÇÃO DE MODELOS PARA A VARIÂNCIA DO PROCESSO

A variação de alguma característica funcional de um produto ou processo é uma preocupação bastante antiga no controle de qualidade do processo. A variação de uma característica de qualidade é avaliada em termos do desvio da média ao valor especificado e, também, em termos da sua própria variância. É falta de qualidade, por exemplo, uma máquina de encher pacotes de café que os enche, em média, com 500g (valor especificado), mas com o peso variando muito de pacote para pacote (variância alta). Esta preocupação com a variabilidade fica muito evidente nos gráficos de controle por variáveis, onde é comum observar conjuntamente um gráfico da média e um gráfico da variância do processo (ver, por exemplo, Paladini, 1995).

Bergman e Hynén (1997) observam que muitos autores, tais como Shewhart, Deming e Taguchi pregam a redução da variabilidade como um meio importante para a melhoria da qualidade. Uma maneira de visualizar esta importância é através da função perda quadrática, tal como definida em Taguchi et al. (1990). Dada uma característica funcional y e o seu valor ideal τ , a função perda mede a qualidade em termos da variação de y em relação a τ . Usando uma expansão em série de Taylor, Taguchi et al. (1990, p. 18) justificam o uso da função perda quadrática, que é definida da seguinte forma:

$$L(y) = k(y - \tau)^2, \quad \text{para } y \in [\tau - \delta, \tau + \delta] \quad (2.22)$$

onde: k é uma constante de proporcionalidade, fazendo com que a função L assuma valores em termos monetários e

$[\tau - \delta, \tau + \delta]$ é o intervalo de especificação da característica y .

Para características de qualidade onde se deseja o maior resultado possível (*maior-é-melhor*) ou menor resultado possível (*menor-é-melhor*), Taguchi propõe expressões alternativas (ver, por exemplo, Taguchi et al., 1990, cap. 2). Contudo, pode-se usar a própria expressão (2.22), considerando τ como um valor suficientemente grande (pequeno), tal que acima (abaixo) dele pode-se admitir perda zero. Neste caso, a função assume valor nulo para y superior (ou inferior) a τ (ver, por exemplo, Pasa et al., 1996). A figura 2.2 ilustra três tipos de função perda.

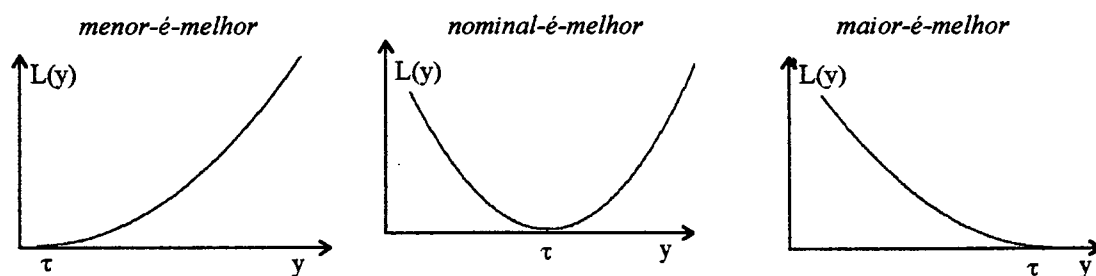


Fig. 2.2 - Ilustração de funções perda quadrática em função do tipo de característica de qualidade.

Num programa de melhoria da qualidade, sendo esta medida em termos da função perda quadrática, torna-se natural o objetivo de minimizar o seu valor esperado. Como visto no capítulo 1, fazendo-se a extensão do domínio de $L(y)$ para todos os reais, tem-se:

$$\mu_L = E\{L(y)\} = k[(\mu_y - \tau)^2 + \sigma_y^2] \quad (2.23)$$

onde μ_y e σ_y^2 são o valor esperado e a variância de y , respectivamente.

Considerando que tanto a média μ quanto a variância σ_y^2 sofrem influência de um vetor $x = (x_1, x_2, \dots, x_k)$ de variáveis do processo, busca-se encontrar o vetor x^0 que minimiza $\sigma_y^2 = \sigma_y^2(x)$ e ajusta $\mu_y = \mu_y(x)$ ao alvo τ . Sarin (1997) afirma que a abordagem de Taguchi focaliza primeiramente a seleção de níveis de fatores controláveis que reduzem a variabilidade. Levar o processo ao alvo é o segundo objetivo. Kackar (1986) argumenta que o ajustamento do valor médio de uma característica de desempenho ao valor alvo é usualmente um problema de engenharia muito mais fácil do que a redução da variação do desempenho.

O objetivo básico é encontrar a combinação de níveis dos fatores que leva à menor perda possível, usando a metodologia de superfície de resposta (RSM). Neste contexto, torna-se necessário caracterizar e identificar os fatores que provocam alterações na média (efeitos de locação) e aqueles que alteram a variância (efeitos de dispersão). Num passo seguinte, busca-se estimar superfícies de resposta para a média e para a variância do processo. E, finalmente, tenta-se encontrar uma solução de compromisso, em termos do ajuste da média ao alvo e da minimização da variância.

A identificação dos efeitos de locação e a estimação da superfície de resposta para a média são problemas exhaustivamente discutidos na literatura (ver seção 2.1). Por outro lado, a identificação dos efeitos de dispersão e a estimação da superfície de resposta para a variância somente ganharam a devida atenção a partir da década de 80, com as idéias de projeto robusto de Taguchi. A dificuldade já começa no planejamento do experimento. Existem projetos experimentais razoavelmente simples para se estimarem os efeitos de locação, mas as suas principais propriedades deixam de valer na presença de efeitos de dispersão (ver Atkinson e Cook, 1995 e Mays e Easter, 1997). Além disso, as variâncias dos efeitos de dispersão costumam ser bastante superiores às variâncias dos efeitos da média, dificultando a identificação e estimação destes efeitos (ver Carroll e Ruppert, 1988).

2.2.1 - Identificação dos efeitos de dispersão

Na maior parte dos trabalhos aplicados, quando se pretende avaliar os efeitos de dispersão, realizam-se mais de um ensaio em cada combinação de níveis dos fatores controláveis, caracterizando um experimento com replicações. Estas replicações podem ser genuínas, onde os ensaios são realizados exatamente sob as mesmas condições, ou, seguindo a abordagem de Taguchi, as repetições podem ser feitas sob variações sistemática de ruídos produzidos em laboratório.

Quando o experimento comporta replicações, o procedimento mais comum para identificar possíveis efeitos de dispersão é calcular, em cada ponto experimental ensaiado com replicação, uma medida de dispersão. Usam-se as medidas de dispersão como se fossem as observações e procede-se a mesma análise usada para identificar efeitos de locação (análise de variância e gráfico normal dos efeitos, por exemplo). As medidas de dispersão mais usadas são as razões *sinai-ruído* e a variância amostral, incluindo transformadas como o desvio padrão e o logaritmo da variância. As razões *sinai-ruído* são as medidas usuais nas técnicas de Taguchi (ver, por exemplo, Phadke, 1989), mas Box (1988) mostra uma série de deficiências destas medidas e sugere, em seu lugar, o uso do logaritmo do desvio padrão ou da variância amostral. A razão para se usar a transformação logarítmica será discutida na seção seguinte.

Box (1988) também discute a questão da métrica da resposta, mostrando a possibilidade de se aplicar uma transformação, onde os efeitos de locação e de dispersão ficam melhores separados, permitindo identificar mais facilmente os efeitos e propiciar modelos aditivos. Em termos empíricos, a transformação é normalmente restrita à família de funções potência, $y^* = y^\lambda$ e estabelecida pelo chamado *gráfico lambda*, onde se observa o nível de separação dos efeitos principais em função de λ . O nível de separação é obtido através das estatísticas usuais da análise de variância (t ou F). Pregibon (ver discussão em Cook e Weisberg, 1989) mostra a possibilidade de interagir o gráfico lambda com o gráfico normal dos efeitos, utilizando-se técnicas de animação.

Box e Meyer (1986a) discutem a questão da identificação dos principais efeitos de dispersão em experimentos fatoriais não replicados de dois níveis (2^k ou 2^{k-r}). Este

trabalho tem grande relevância prática, pois é muito comum iniciar-se com experimentos preliminares, constituídos de muitos fatores, onde o uso de replicações pode tornar o experimento demasiadamente caro. A técnica, baseada no princípio da esparsidade, consiste em avaliar primeiramente os efeitos de locação. Depois, subtraem-se dos dados os principais efeitos de locação, produzindo-se resíduos, os quais trazem informações sobre a dispersão. Calculam-se, a partir dos resíduos, os resultados da seguinte estatística:

$$F_i^* = \log \left[\frac{S^2(i^+)}{S^2(i^-)} \right], \quad i = 1, 2, \dots, n-1 \quad (2.24)$$

onde: n é o número de ensaios;

$S^2(i^+)$ é a variância dos resíduos, considerando as observações feitas no nível mais alto do efeito i e

$S^2(i^-)$ é a variância dos resíduos, considerando as observações feitas no nível mais baixo do efeito i ;

Se o efeito i ($i = 1, 2, \dots, n-1$) é nulo, então F_i^* tem distribuição aproximadamente normal padrão. Montgomery (1997, p. 335) sugere identificar os efeitos não negligenciáveis através do gráfico normal dos efeitos. Testes estatísticos formais não são recomendados, pois a distribuição dos F_i^* ($i = 1, 2, \dots, n-1$) não é exata e depende do número dos efeitos considerados no cálculo dos resíduos. Uma estatística alternativa pode ser vista em Bergman e Hynén (1997), a qual é independente dos efeitos de locação usados no cálculo dos resíduos e que tem distribuição exata conhecida, o que permite avaliar a significância dos efeitos através de testes estatísticos. Contudo, os mesmos autores alertam para o fato dos testes não serem independentes entre si e, além disso, os efeitos de dispersão podem estar confundidos com interações e, assim, viciar os resultados dos testes.

Nair e Pregibon (1988) generalizam a estatística de Box e Meyer para experimentos fatoriais de dois níveis replicados. Observam que esta estatística normalmente é viciada e, através de um estudo de simulação do tipo Monte Carlo, verificam que se houver mais de um efeito de dispersão, a utilização da estatística de Box e Meyer, como instrumento de identificação, pode levar a conclusões incorretas.

Por outro lado, quando se tem apenas um efeito de dispersão, o estudo mostra que a estatística de Box e Meyer é eficiente.

Myers e Montgomery (1995, p. 526) argumentam que a variância de replicação é mais confiável como informação no processo da variância do que aquela obtida a partir dos resíduos. No último caso, a confiabilidade da informação da variância depende, em grande parte, da qualidade do modelo da média que gerou os resíduos.

Nas seções seguintes, a ênfase maior está orientada para a estimação dos coeficientes dos modelos. Contudo, os procedimentos também podem ser usados para se identificarem os efeitos de dispersão, usando-se os testes estatísticos usuais da análise de regressão ou abordagens gráficas.

2.2.2 - O modelo básico

Na seção 2.1.2, foi abordada a questão da construção da superfície de resposta para o valor esperado, $E(y/x) = f(x, \beta)$, $x \in D \subset \mathcal{R}^k$. Para realizar o projeto de parâmetros (projeto robusto) através da RSM, torna-se fundamental o conhecimento da variabilidade da resposta em função dos fatores controláveis, donde uma superfície de resposta para a variância, $\text{Var}(y/x) = h(x, \beta)$, $x \in D \subset \mathcal{R}^k$, tem grande importância. Este tema será abordado nesta e nas próximas seções.

Assume-se que a qualidade do produto ou processo possa ser avaliada por alguma característica funcional, y , caracterizada como uma variável aleatória contínua. Um experimento é realizado sob M diferentes pontos experimentais, formado por combinações de níveis dos fatores controláveis. Em cada um dos M pontos experimentais, y é observado m_i vezes ($i = 1, 2, \dots, M$), totalizando

$n = \sum_{i=1}^M m_i$ observações. Para facilitar a notação, usar-se-á $m_i = m$ ($i = 1, 2, \dots, M$),

donde pode-se dizer que o experimento é realizado em M pontos experimentais, com m replicações e, portanto, com número total de observações $n = Mm$. Em termos práticos, contudo, pode-se usar diferentes números de observações nos pontos experimentais ensaiados.

A média de y pode estar relacionada com um conjunto de fatores controláveis, $x = (x_1, x_2, \dots, x_k)'$ e a variância de y com um conjunto de fatores controláveis, $z = (z_1, z_2, \dots, z_t)'$. O vetor z pode ser o próprio vetor x . Assumindo a possibilidade de um modelo linear ajustar-se aos dados, tem-se:

$$y_{ij} = x_{ij}'\beta + \sigma_i e_{ij}, \quad (2.25)$$

$$g(\sigma_i^2) = z_i'\theta, \quad i = 1, 2, \dots, M \quad \text{e} \quad j = 1, 2, \dots, m \quad (2.26)$$

onde: y_{ij} são os dados oriundos do experimento ou os resultados de alguma transformação sobre os dados;

x_{ij} são os vetores formados pelas linhas da matriz de planejamento X do modelo de regressão da resposta (ou da média);

β é o vetor de parâmetros do modelo da resposta (ou da média);

σ_i^2 são as variâncias nos M pontos experimentais;

g é alguma função que facilite a colocação de um modelo linear para a variância;

z_i são os vetores formados pelas linhas da matriz de planejamento Z do modelo de regressão da variância;⁵

θ é o vetor de parâmetros do modelo da variância;

e_{ij} são os erros aleatórios, que supostamente têm valores esperados nulos, variâncias unitárias e, preferencialmente, distribuições normais;

M é o número de pontos experimentais ensaiados e

m é o número de observações em cada ponto experimental (número de replicações).

Observa-se que o modelo tradicional de regressão considera somente a equação (2.25), assumindo $\sigma_i = \sigma$ constante (ver seção 2.1.2). O modelo (2.25) será chamado de *modelo da resposta* (ou *modelo da média*, quando visto em termos de seu valor esperado) e o modelo (2.26) de *modelo da variância*. O primeiro objetivo é estimar os vetores de parâmetros β e θ , referentes aos modelos da média e da variância, respectivamente.

⁵ Está se considerando Z com M linhas, mas pode-se também utilizar Z com $n = Mm$ linhas sob a forma z_{ij} e, neste caso, as variâncias também teriam a forma σ_{ij}^2 .

A estimação do vetor de parâmetros β , referente ao modelo (2.25), pode ser feita pelas técnicas usuais de regressão, aplicando-se o método dos mínimos quadrados sobre as observações, y_{ij} ($i = 1, 2, \dots, M$ e $j = 1, 2, \dots, m$), ou sobre as médias, \bar{y}_i ($i = 1, 2, \dots, M$), calculadas nos diferentes pontos experimentais ensaiados. Desde que se tenha o mesmo número de observações em cada ponto experimental, ambos os procedimentos fornecem as mesmas estimativas pontuais. Nair e Pregibon (1986), Box (1988) e Guedes (1996) enfatizam a necessidade de se buscar uma transformação em \bar{y}_i (ou em y_{ij}) para estabilizar a variância. Uma abordagem alternativa é a utilização da técnica de mínimos quadrados generalizados (MQG), que será discutida na seção seguinte.

A ênfase maior, nesta seção, é a estimação do vetor de parâmetros θ , referente ao modelo da variância (2.26). Um problema a se considerar é que ao se estabelecer a equação de regressão da variância, o termo de erro deste modelo tem normalmente variância grande, pois ela é função do quarto momento da resposta, que corresponde, aproximadamente, a duas vezes o quadrado da variância. Em consequência, as estimativas dos parâmetros do modelo da variância tornam-se menos precisas do que as estimativas dos parâmetros do modelo da resposta. Da mesma forma, o poder dos testes para detectar efeitos significativos é bem menor no modelo da variância do que no modelo da resposta.

2.2.3 - Modelos baseados em variâncias amostrais

Grande parte dos trabalhos práticos de projeto de parâmetros, que envolvem modelos de variâncias, usam como informação básica a variância amostral, calculada em cada ponto experimental ensaiado, ou seja:

$$S_i^2 = \frac{\sum_{j=1}^m (y_{ij} - \bar{y}_i)^2}{m-1}, \quad i = 1, 2, \dots, M. \quad (2.27)$$

Uma característica comum dos modelos que usam a variância (ou o desvio padrão) amostral é que o processo de estimação de θ independe do modelo da média. E este procedimento será referenciado, neste trabalho, de método S2.

Mesmo considerando os dados normalmente distribuídos, não existe uma forma padrão para se estabelecer a equação de regressão para estimar os parâmetros do modelo da variância. Segue alguns exemplos, onde θ denota o vetor de parâmetros e u_i o termo aleatório ($i = 1, 2, \dots, M$), sempre correspondente ao modelo considerado:

$$S_i^2 = z_i' \theta + u_i \quad (\text{Caten, 1995})$$

$$S_i = z_i' \theta + u_i \quad (\text{Vining e Myers, 1990})$$

$$\log(S_i) = z_i' \theta + u_i \quad (\text{Box, 1988})$$

$$\log(S_i^2) = z_i' \theta + u_i \quad (\text{Myers e Montgomery, 1995})$$

onde a denotação “log” representa o logaritmo natural ou neperiano.

Na maioria dos trabalhos práticos, incluindo os quatro citados, o processo de estimação do vetor de parâmetros θ é realizado pelo método dos mínimos quadrados ordinários (MQO).

Myers e Montgomery (1995, p. 508) justificam a transformação logarítmica, aplicada à variância amostral, por resultar num modelo de regressão que satisfaz aproximadamente os pressupostos de um modelo linear. Além disso, esta transformação tende a gerar modelos mais simples, reduzindo os efeitos de curvaturas e de interações. Outra vantagem no uso do logaritmo, apontada por Carroll e Ruppert (1988, p. 65), é que sempre garante predições positivas para as variâncias.

A base teórica da transformação logarítmica, $\log(S_i^2)$ ou $\log(S_i)$, é descrita em Bartlett e Kendall (1946). Estes autores mostram que esta transformação estabiliza a variância dos erros aleatórios, isto é, no modelo com logaritmo tem-se $\text{Var}(u_1) = \text{Var}(u_2) = \dots = \text{Var}(u_M)$, condição necessária para que o método de estimação MQO produza estimadores eficientes. Porém, os mesmos autores alertam que a distribuição dos erros só será aproximadamente normal se o número de replicações, m , for razoavelmente grande, sugerindo a aplicação do método quando $m > 10$ ou, com muita cautela, quando $m > 5$. A seguir são apresentados alguns dados que justificam a preocupação dos autores.

A figura 2.3 apresenta algumas características da distribuição do termo de erro de $\log(S^2)$, em função do número de observações, m , usadas no cálculo de S^2 . Estes resultados foram extraídos da tabela I de Bartlett e Kendall (1946) e pressupõe distribuição normal para a resposta y .

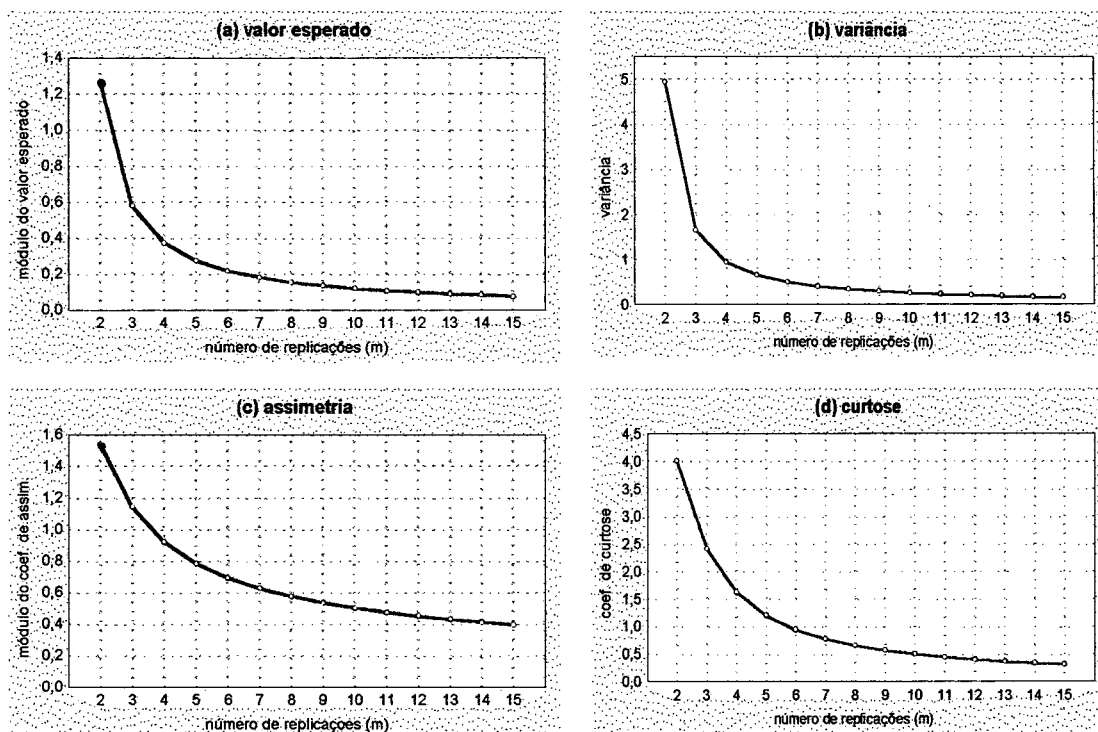


Fig. 2.3 - As principais características da distribuição do termo de erro de $\log(S^2)$, em função do número de replicações. Os pontos em destaque em (a) e (c) são negativos.

Observa-se, na figura 2.3 (a), que o valor esperado do termo de erro não é nulo, o que acarreta, segundo Harvey (1976), inconsistência na estimação do termo constante da equação (2.26).⁶ Esta inconsistência pode ser corrigida adicionando-se uma certa quantidade no estimador. Conforme mostra a figura 2.3 (a), para $m = 2, 3, 4, 5$ e 6 , pode-se adicionar no estimador do termo constante os valores iguais a 1,270, -0,577, -0,369, -0,270 e -0,213, respectivamente.

Na figura 2.3(b), verifica-se que a variância do termo de erro é reduzida rapidamente com o aumento das replicações, o que leva a aumentar a precisão das estimativas. Em (c) e (d), observa-se que, para m pequeno, a distribuição não pode ser considerada aproximadamente normal, o que pode invalidar os testes estatísticos t e F , usualmente utilizados na análise de regressão.

⁶ De maneira informal, pode-se dizer que um estimador é *consistente* quando ele tende para o verdadeiro valor do parâmetro quando o tamanho da amostra cresce infinitamente.

Uma característica interessante nos dados de Bartlett e Kendall é que para $m = 2$ a assimetria é negativa (igual a $-1,535$). Isto deve ocorrer devido à possibilidade de haver valores de S^2 próximos de zero. A aplicação do logaritmo nestes valores leva a resultados extremamente pequenos, provocando artificialmente pontos discrepantes (*inlier*, segundo a denominação de alguns autores, como Davidian e Carroll, 1987). Para $m > 2$, a probabilidade de ocorrências de *inliers* torna-se menor. Outro argumento do inconveniente de se usar o método S2 para m pequeno é ilustrado na figura 2.4.

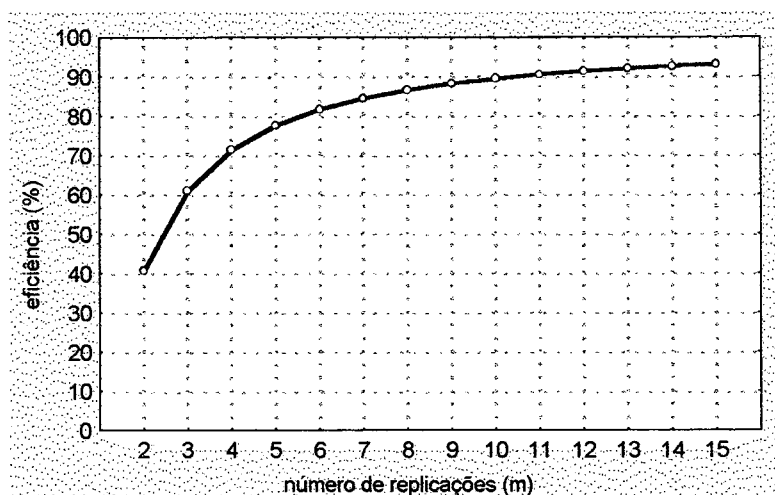


Fig. 2.4 - Eficiência de $\log(S^2)$ em relação a S^2 na estimação do parâmetro σ^2 da distribuição normal. Dados extraídos de Bartlett e Kendall (1946).

A figura 2.4 ilustra que, para se estimar a variabilidade de uma distribuição normal, o uso de $\log(S^2)$ é bem menos eficiente do que a utilização direta de S^2 , quando o número de replicações é pequeno. Os dados desta relação são obtidos em termos assintóticos, isto é, quando o número de pontos experimentais cresce infinitamente (ver também, Harvey, 1976 e Davidian e Carroll, 1987). Contudo, conforme argumentam Engel e Huele (1996), a eficiência, vista em termos assintóticos, pode ser muito diferente para amostras pequenas ou moderadas, como é usual em dados experimentais.

Nair e Pregibon (1988) comparam alguns métodos para identificar efeitos de variância em experimentos fatoriais do tipo 2^k com replicações. Baseados em algumas deduções matemáticas e em estudos de simulação do tipo Monte Carlo, eles observam que o método S2 funciona razoavelmente bem como instrumento de

identificação, mas pode ser ineficiente quando existem somente alguns efeitos não nulos, particularmente em grandes experimentos. Em tais casos, sugerem o método da máxima verossimilhança (ver seção 2.2.5).

Na literatura, encontram-se outras transformações da variância amostral, mas sem referencial teórico. Por exemplo, Vining e Myers (1990) e Castilho (1996) usam a transformação raiz quadrada sobre as variâncias amostrais, gerando um modelo para o desvio padrão do processo. Contudo, se houver interesse num modelo para a variância propriamente dita, que é importante quando se usa a abordagem da função perda quadrática, necessita-se aplicar a transformação inversa. A transformação inversa da raiz quadrada (função quadrática) leva a uma expressão em que os fatores de entrada do processo não se mantêm em termos isolados, dificultando a sua aplicabilidade. Por outro lado, a transformação logarítmica gera um modelo exponencial para a variância, transformando as somas em produtos, cuja interpretação não traz maiores dificuldades práticas.

2.2.4 - Modelos baseados nos resíduos

Nesta seção, o modelo de regressão da variância será estabelecido a partir dos resíduos do modelo da média, o que será referenciado de método RQ. Formalmente, aplica-se o método MQO sobre os dados (ou sobre os resultados de alguma transformação dos dados), que supostamente satisfazem o seguinte modelo:

$$y_i = x_i' \beta + \sigma_i \varepsilon, \quad i = 1, 2, \dots, n, \text{ sendo } n = Mm. \quad (2.28)$$

As estimativas para o vetor de médias, $\hat{\mu} = \{\mu_i\}$, são obtidas por

$$\hat{\mu} = X\hat{\beta} \quad (2.29)$$

onde $\hat{\beta}$ é o estimador de MQO de β , calculado por

$$\hat{\beta} = (X'X)^{-1}(X'y) \quad (2.30)$$

Com a estimativa do vetor de médias, pode-se obter os *resíduos brutos*: $y_i - \hat{\mu}_i$ ($i = 1, 2, \dots, n$). Carroll e Ruppert (1988, p. 78) sugerem levar em conta a perda de graus de liberdade na estimação do modelo da média, fazendo-se o cálculo dos *resíduos corrigidos* da seguinte maneira:

$$\hat{e}_i = \frac{y_i - \hat{\mu}_i}{\sqrt{1 - h_i}} \quad (2.31)$$

onde $h_i = x_i(X'X)^{-1}x_i'$, sendo x_i a i -ésima linha da matriz de planejamento X . A divisão por $\sqrt{1 - h_i}$ tornam os resíduos quadráticos, \hat{e}_i^2 ($i = 1, 2, \dots, n$), estimadores aproximadamente não viciados das correspondentes variâncias, σ_i^2 (ver Davidian e Carroll, 1987).

Os programas computacionais que realizam análise de regressão normalmente calculam os chamados *studentized residuals*, que são os resíduos corrigidos divididos pela raiz quadrada do resíduo quadrático médio. Logo, a equação de regressão da variância também pode ser estabelecida em termos dos *studentized residuals*.⁷

O uso dos resíduos quadráticos como base para estabelecer modelos de variância tem uma literatura mais abrangente. Historicamente, o método RQ foi tratado como um método de testar e corrigir o problema da variância não constante (heterocedasticidade), na estimação do modelo da média.

A equação de regressão da variância não tem sido apresentada sob a mesma forma funcional. Por exemplo, Amemiya (1977) e Mandy e Martins-Filho (1993) consideram a heterocedasticidade aditiva, isto é, supõem o seguinte modelo:

$$\hat{e}_i^2 = z_i'\theta + u_i \quad i = 1, 2, \dots, n \quad (2.32)$$

enquanto Harvey (1976) e Aitkin (1987) adotam a heterocedasticidade multiplicativa, ou seja:

$$\hat{e}_i^2 = \exp\{z_i'\theta + u_i\}, \quad (2.33)$$

$$\text{ou} \quad \log(\hat{e}_i^2) = z_i'\theta + u_i, \quad i = 1, 2, \dots, n. \quad (2.34)$$

Como a transformação logarítmica estabiliza a variância dos u_i ($i = 1, 2, \dots, n$), a estimação de θ , na equação (2.34), pode ser feita por MQO. Modelos do tipo (2.34) aparecem em muitos estudos experimentais, como, por exemplo, em Box e Meyer (1986a), Nair e Pregibon (1988), Myers e Montgomery (1995) e Nelder e Lee (1998).

⁷ Manteve-se a expressão "*studentized residuals*" escrita na língua inglesa por não se encontrar uma tradução adequada.

Carroll e Rupert (1988) discutem o problema da estimação em um modelo bastante genérico de regressão, como segue:

$$\begin{aligned} E(y_i) &= f(x_i, \beta) \quad e \\ \text{Var}(y_i) &= \sigma^2 g^2\{\mu_i(\beta), z_i, \theta\} \end{aligned} \quad (2.35)$$

O modelo (2.35), além de abranger os dois casos anteriores, permite incluir modelos não lineares e modelos em que a variância depende funcionalmente da média. Chan e Mark (1995) apresentam uma metodologia para se obter modelos para a variância do processo, considerando explicitamente que a variância é função da média e dos fatores de entrada. O trabalho de Chan e Mark enfatiza o uso do método RQ como uma alternativa de se obter modelos para a variância do processo, em experimentos não replicados.

Quando houver evidência de que as variâncias das observações não são constantes, pode-se refazer as estimativas do modelo da média (ou da resposta) pelo método conhecido como mínimos quadrados generalizados (MQG), que deve gerar estimadores mais eficientes do que o tradicional método dos mínimos quadrados ordinários (MQO). No método MQG, atribuem-se diferentes pesos às observações de acordo com o inverso de suas variâncias. Mais especificamente, busca-se o vetor $\hat{\beta}$ que minimiza

$$\sum_{i=1}^n w_i (y_i - x_i' \beta)^2, \quad \text{onde } w_i = \frac{1}{\sigma_i^2} \quad (2.36)$$

Na prática, as σ_i^2 ($i = 1, 2, \dots, n$) são desconhecidas e, por isto, usam-se as estimativas produzidas pelo modelo da variância.⁸

A eficiência do MQG em relação ao MQO é grande quando também forem grandes as diferenças entre as variâncias das observações. Por outro lado, o MQG pode ser menos eficiente do que o MQO se as diferenças entre as variâncias forem pequenas. Em consequência, o MQG só deve ser usado após testes específicos de

⁸ Em termos computacionais, o método MQG pode ser feito por uma rotina de MQO. Antes de proceder as estimações, divide-se as observações da variável dependente e de todos os fatores (inclusive o vetor de elementos unitários) pelo inverso do desvio padrão da correspondente observação. A aplicação de MQO nestas variáveis transformadas equivale ao MQG nas variáveis originais.

heterocedasticidade, como por exemplo os descritos em White (1980) ou em Judge et al. (1988, seção 9.3.5). Observa-se que o próprio teste F do modelo da variância pode ser considerado como um teste de heterocedasticidade.

Quando há evidência de variância não constante, Engel e Huele (1996) sugerem o uso de MQG, argumentando que nem sempre é possível encontrar uma transformação adequada para estabilizar a variância das observações. E, no método MQG, conforme apontam Carroll e Cline (1988) e Chen e Shao (1993), a estimação dos pesos, w_i ($i = 1, 2, \dots, n$), torna-se mais eficiente quando se usam resíduos quadráticos (método RQ) do que quando se usam variâncias amostrais (método S2).

No método RQ, como o modelo da variância depende da adequação do modelo da média e, sabendo-se que a estimativa deste pode ser melhorada com a correção da heterocedasticidade, sugere-se o seguinte procedimento iterativo para o modelo multiplicativo (ver, por exemplo, Engel e Huele, 1996):

- a) calculam-se estimativas iniciais para os parâmetros do modelo da média usando MQO sobre os dados do experimento;
- b) calculam-se os resíduos corrigidos: \hat{e}_i ($i = 1, 2, \dots, n$);⁹
- c) estabelece-se a equação de regressão em termos do logaritmo dos resíduos quadráticos, $\log(\hat{e}_i^2)$ e aplica-se MQO produzindo o modelo estimado $\hat{\sigma}^2 = \exp\{Z\hat{\theta}\} + 1,27$, onde o valor 1,27 é sugerido por Harvey (1976) para corrigir o vício na estimativa do termo constante;
- d) calculam-se as variâncias preditas pelo modelo de (c) e
- e) aplica-se MQG no modelo da média, usando $w_i = \frac{1}{\hat{\sigma}_i^2}$ como peso das observações.

Repete-se os passos de (b) a (e) pelo menos mais uma vez.

⁹ Observa-se que a maioria dos trabalhos que se utilizam dos resíduos para estabelecer modelos para a variância não usam a correção sugerida na expressão (2.31). Nos projetos fatoriais do tipo 2^k , a correção é constante para todos os resíduos e, portanto, só altera a estimativa do termo constante do modelo.

A utilização da transformação logarítmica sobre os resíduos quadráticos pode ser problemática quando houver resíduos nulos ou quase nulos. Carroll e Ruppert (1988) e Engel e Huele (1996) orientam para excluir tais valores. Ferrer e Homero (1995) sugerem, para os resíduos nulos, somar a quantidade $(0,5)\min|r_i|$, onde $|r_i|$ representa o valor absoluto dos resíduos não nulos.

Ferrer e Homero (1993) mostram, através de simulações, que as estimativas dos coeficientes do modelo de $\log(\hat{\epsilon}_i^2)$ são ligeiramente viciadas quando a amostra é pequena, tanto para resíduos obtidos por MQO quanto para resíduos obtidos por MQG. Além disso, o vício apresenta-se com sinal invertido. Com resíduos provindos de MQO, tende haver subestimação no valor absoluto dos coeficientes, enquanto que os resíduos provindos de MQG tendem a superestimar a magnitude dos coeficientes.

No modelo aditivo, para se usar MQG iterativo, pode-se adotar $w_i = \frac{1}{\hat{\sigma}_i^4}$ como pesos dos resíduos quadráticos, pois, para dados normalmente distribuídos, $\text{var}(\hat{\epsilon}_i^2) \approx 2\sigma_i^4$ (ver Carroll e Ruppert, 1988, p. 78). Neste caso, o algoritmo de estimação pode ser descrito da seguinte forma:¹⁰

- a) calculam-se estimativas iniciais para os parâmetros do modelo da resposta usando MQO sobre os dados do experimento;
- b) calculam-se os resíduos corrigidos: $\hat{\epsilon}_i$ ($i = 1, 2, \dots, n$);
- c) consideram-se os resíduos quadráticos $\hat{\epsilon}_i^2$ como se fossem observações de um modelo de regressão e calculam-se estimativas iniciais para os parâmetros do modelo da variância por MQO;
- d) aplica-se MQG no modelo da média, usando $w_i = \frac{1}{\hat{\sigma}_i^2}$ como peso das observações, onde $\hat{\sigma}_i^2$ é obtido na estimação do modelo da variância (passo anterior);

¹⁰ Este algoritmo é uma simplificação do apresentado em Carroll e Ruppert (1988, p. 69).

e) calculam-se novamente os resíduos quadráticos e aplica-se MQG no modelo das variâncias, usando-se como peso das observações $w'_i = \frac{1}{\hat{\sigma}_i^4}$.

Os passos (d) e (e) podem ser repetidos várias vezes (processo iterativo). Carroll e Ruppert (1988) sugerem pelo menos duas iterações, o que corresponde em repetir (d) e (e) ao menos mais uma vez. Uma desvantagem do modelo aditivo é apontada por Raj (1975), que observa que o modelo pode gerar previsões negativas quando a amostra é pequena e a heterocedasticidade é forte.

Um aspecto curioso no MQG iterativo é que nem sempre um maior número de iterações gera melhores estimativas. Uma justificativa para isto é dada em Hooper (1993), que afirma que algumas variâncias podem ter como estimativas valores muito pequenos, gerando grandíssimos pesos e, assim, acarretando estimativas piores nas iterações seguintes (ver, também, Ferrer e Homero, 1993).

Carroll, Wu e Ruppert (1988) mostram que, a partir da segunda iteração, as variâncias das estimativas permanecem aproximadamente constantes, desde que a variância do processo independa da média e os erros tenham distribuição simétrica. Engel e Huele (1996), baseados num estudo de simulação realizado sobre um planejamento fatorial completo 2^5 , mostram que duas iterações apresentam melhores resultados. A simulação de Engel e Huele mostra, também, que o uso de iterações melhoram as estimativas quando a heterogeneidade é forte, mas podem piorar quando a diferença entre as variâncias não for muito grande.

Carroll e Ruppert (1988, p. 80) mostram a possibilidade de se usarem os resíduos absolutos no lugar dos resíduos quadráticos. Em termos do processo de estimação MQG iterativo, esta substituição leva a uma perda de eficiência em torno de 12% para dados normalmente distribuídos, porém a eficiência relativa melhora rapidamente na presença de valores discrepantes. Davidian e Carroll (1987) observam a possibilidade de se utilizarem outras transformações, tais como potências $\frac{1}{2}$ (raiz quadrada) ou $\frac{2}{3}$, que tendem a produzir resultados com distribuições mais próximas da normal.

Davidian e Carroll (1987) desenvolvem uma teoria assintótica para comparar a eficiência no uso da variância amostral (método S2) em relação ao uso do resíduo

quadrático (método RQ), sob o método de estimação MQG iterativo.¹¹ Com dados normalmente distribuídos, a eficiência relativa de S2 em relação à RQ é dada por $(m-1)/m$, onde m é o número de replicações.¹² Logo, se houver apenas duas observações em cada condição experimental, a aplicação de S2 leva a uma perda de 50% na eficiência. Por outro lado, quando m é grande, o uso de variâncias amostrais tem eficiência quase semelhante ao uso de resíduos quadráticos. Outras comparações entre estes métodos serão discutidas no capítulo 3.

2.2.5 - Outros métodos de estimação

Embora este trabalho enfatiza a utilização do método dos mínimos quadrados, por ser mais simples e mais conhecido na engenharia, esta seção apresenta algumas propostas de estimação conjunta dos parâmetros dos modelos da média e da variância, baseadas no método da máxima verossimilhança (MV) e suas variantes.

O método MV parte da suposição de que os dados seguem uma determinada distribuição de probabilidade. Os estimadores (EMV) são identificados com os valores dos parâmetros que tornam a distribuição hipotética tão mais verossímil possível com respeito aos dados observados (ver detalhes em Cordeiro, 1992).

Harvey (1976) apresenta a função de verossimilhança conjunta para os vetores de parâmetros β e θ , relativos aos modelos da média e da variância, expressões (2.25) e (2.26), respectivamente. Em termos do modelo multiplicativo e com a suposição de normalidade dos dados, ela é dada por

$$\log L = \frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n (z_i' \theta) - \frac{1}{2} \sum_{i=1}^n \left[(y_i - x_i' \beta)^2 e^{-z_i' \theta} \right] \quad (2.37)$$

Seguindo o método do escore (ver Cordeiro, 1992, p. 9), a obtenção dos estimadores dos vetores β e θ , através da maximização de $\log L$, é feita a partir de soluções iniciais, $\hat{\beta}^{(0)}$ e $\hat{\theta}^{(0)}$ (obtidas por MQO, por exemplo), pelo seguinte procedimento iterativo:

¹¹ Na teoria assintótica, supõe-se que o tamanho da amostra tenda para infinito e, portanto, pode ser considerada aproximadamente válida para as situações em que se tem amostras grandes. No presente contexto, em que $n = Mm$, considera-se m fixo e M tendendo para infinito.

¹² O conceito de eficiência relativa entre dois estimadores é apresentado na seção 3.1.

$$\hat{\beta}^{(j+1)} = \hat{\beta}^{(j)} + \left(\sum_i e^{-z_i' \hat{\theta}^{(j)} x_i x_i'} \right)^{-1} \sum_i x_i e^{-z_i' \hat{\theta}^{(j)}} (y_i - x_i' \hat{\beta}^{(j)}) \quad (2.38)$$

$$\hat{\theta}^{(j+1)} = \hat{\theta}^{(j)} + \left(\sum_i z_i z_i' \right)^{-1} \sum_i z_i \left[e^{-z_i' \hat{\theta}^{(j)}} (y_i - x_i' \hat{\beta}^{(j)})^2 - 1 \right] \quad (2.39)$$

Harvey (1976) compara, em termos assintóticos, o EMV com o estimador de MQO com transformação logarítmica, na estimação de θ . Observa que o EMV é bem mais eficiente, pois o termo de escala da variância do EMV é igual a 2, enquanto a variância estabilizada com a transformação logarítmica é igual a 4,9348. Por outro lado, Ferrer e Homero (1993) mostram, através de estudos de simulação do tipo Monte Carlo, que o método MV pode ser inferior ao MQO com transformação logarítmica em amostras pequenas. Por exemplo, num problema de regressão simples, com amostra de tamanho 11, a estimativa do modelo da variância por MV apresentou como variância do erro o valor 6,7 e, portanto, superior ao valor estabilizado pela transformação logarítmica (4,9348). Engel e Huele (1996) fazem comentários análogos, porém não mostram os resultados.

Carroll e Ruppert (1982a) observam que os EMV são muito sensíveis (ou pouco robustos) à presença de valores discrepantes e à má especificação dos modelos funcionais da média e da variância. Em relação à presença de valores discrepantes, Carroll e Ruppert (1982b) e Giltman et al. (1986) sugerem MV ou MQG acompanhados de técnicas de estimação robustas. Observa-se, contudo, que o processo de estimação torna-se demasiadamente complexo neste caso.

Vários autores, tais como Harville (1977), Carroll e Ruppert (1988) e Verbyla (1993), sugerem embutir na equação (3.39) a perda dos graus de liberdade na estimação do modelo da média. Este método costuma ser referenciado como máxima verossimilhança restrito e produz estimativas mais precisas do que o método MV, especialmente quando o número de parâmetros é relativamente grande em relação ao número de ensaios do experimento.

Aitkin (1987), Gordon e Smyth (1989), Grego (1993) e Nelder e Lee (1998) enfatizam a utilização da metodologia dos modelos lineares generalizados (GLM) no processo de estimação conjunta dos parâmetros dos modelos da média e da variância do processo. Esta metodologia permite fazer estimativas por MV (ou MV restrita)

para uma ampla família de distribuições de probabilidade, embora o caso em que a resposta tenha distribuição normal é citado com maior ênfase.

2.2.6 - Modelos com efeitos de locação e de dispersão

Vários autores, dentre eles Welch et al. (1990), Shoemaker et al. (1991) e Lucas (1994), fazem uso de um modelo misto para as observações considerando os fatores controláveis, x_1, x_2, \dots, x_k , como fixos e os fatores de ruído, u_1, u_2, \dots, u_L , como aleatórios. Considera-se que os principais fatores de ruído que afetam a qualidade, tais como variações não controláveis do processo, fatores ambientais e diferentes condições de uso, possam ser simulados em laboratório durante a realização do experimento. Modelando adequadamente as observações, consegue-se gerar modelos para a média e variância do processo, a partir do ajuste de uma única equação de regressão por MQO.

Para a aplicação deste método, torna-se necessário um bom conhecimento sobre os fatores de ruído, pois considera-se que toda a parte estrutural da variância esteja contemplada nos fatores de ruído incluídos no modelo. Além disso, precisa-se conhecer a média e a variância de cada fator de ruído.

Seja o seguinte modelo, que envolve efeitos lineares dos fatores de controle e de ruído, efeitos quadráticos dos fatores de controle e interações entre fatores de controle e de ruído:

$$y = \beta_0 + x' \beta_1 + x' \beta_2 x + u' \xi + x' \phi u + \varepsilon \quad (2.38)$$

onde: y é a variável resposta;

β_0 é o termo constante;

x é o vetor de fatores controláveis;

β_1 é o vetor de coeficientes dos termos lineares dos fatores controláveis;

β_2 é a matriz de coeficientes dos termos quadráticos dos fatores controláveis, sendo que a diagonal principal é formada pelos coeficientes dos termos quadráticos puros e fora da diagonal principal, pelos coeficientes dos termos de interação divididos por dois (análoga à matriz B da expressão 2.19);

u é o vetor de fatores de ruído, onde supõe-se que cada elemento tenha média zero e variância conhecida;

- ξ é o vetor de coeficientes dos termos lineares dos fatores de ruído;
 ϕ é a matriz de coeficientes dos termos de interação entre fatores controláveis e de ruído e
 ε é o termo de erro, assumido com média zero e variância σ^2 .

Se conveniente, pode-se também incluir termos quadráticos para os fatores de ruído e interações entre fatores controláveis e entre fatores de ruído. Contudo, conforme pondera Myers e Montgomery (1995, p. 492), o modelo apresentado satisfaz a maioria das aplicações práticas. Aplicando-se o operador de valor esperado e de variância em (2.38), tem-se os seguintes modelos para a média e para a variância do processo:

$$\mu_y = \beta_0 + x'\beta_1 + x'\beta_2 x \quad (2.39)$$

$$\sigma_y^2 = (\xi' + x'\phi)(\xi' + x'\phi)' \sigma_u^2 + \sigma^2 = \left(\frac{\partial y}{\partial u} \right)' \left(\frac{\partial y}{\partial u} \right) \sigma_u^2 + \sigma^2 \quad (2.40)$$

onde: σ_u^2 é o vetor de variâncias dos fatores de ruído e

σ^2 é a variância do termo de erro.

Observa-se na expressão (2.40) que o vetor de parâmetros δ caracteriza a relação da variância do processo com o vetor x de fatores controláveis. Isto significa, por exemplo, que, se $\phi = 0$, então a variância não depende de x e, portanto, não se tem um problema de projeto robusto.

A estimação dos parâmetros do modelo (2.36) pode ser feita por MQO. Seja a seguinte equação de regressão obtida após a aplicação de MQO:

$$\hat{y} = b_0 + x'b + x'Bx + u'c + x'du \quad (2.41)$$

Ao contrário dos casos anteriores, onde o modelo estimado da resposta coincidia com o modelo estimado da média do processo, a equação (2.41) agrega os dois modelos. Eles podem ser escritos através da substituição dos parâmetros das equações (2.39) e (2.40) pelas suas respectivas estimativas, dadas na equação (2.41), resultando em:

$$\hat{\mu}_y = b_0 + x'b + x'Bx \quad (2.42)$$

$$\hat{\sigma}_y^2 = (c' + x'd)(c' + x'd)' \sigma_u^2 + \hat{\sigma}^2 \quad (2.43)$$

onde $\hat{\sigma}^2$ é o resíduo quadrático médio da equação de regressão (2.41).

Shoemaker et al. (1991) comentam que a maior vantagem de se obterem as equações da média e da variância por este processo é a possibilidade de se fazerem experimentos com menor número de ensaios, pois pode-se *combinar* fatores de controle com fatores de ruído, ao contrário do processo usual, onde necessita-se *cruzar* todos os fatores de controle com todos os fatores de ruído. Lucas (1995) apresenta um quadro comparativo do número de ensaios necessários em cada abordagem, para diferentes projetos de experimentos. Fica evidente no artigo de Lucas que o número de ensaios, quando se usa a presente metodologia, é bem menor do que na forma clássica. Rosenbaum (1996) discute aspectos teóricos da combinação entre fatores de controle e de ruído, mostrando também como construir tais projetos experimentais a partir de confundimentos de projetos fatoriais fracionados, que são usualmente tabelados em livros textos.

Uma desvantagem da abordagem apresentada neste tópico, também apontada por Shoemaker et al. (1991), é que ela depende muito fortemente das suposições sobre o modelo. Se algum fator de ruído relevante for excluído do modelo, as equações deixam de representar adequadamente os processos. Além disso, é necessário que os engenheiros do processo conheçam as médias e variâncias dos fatores de ruído ou se faça um estudo observacional para estimá-los.

Engel e Huele (1996) sugerem substituir $\hat{\sigma}^2$ na equação (2.41) por uma função dos fatores do processo x_1, x_2, \dots, x_k , obtida por um dos métodos anteriormente discutidos. Desta forma, a variância do processo seria formada pela agregação de duas equações. A primeira sendo constituída do primeiro termo de (2.41), a qual modela os efeitos dos ruídos produzidos no laboratório, e a segunda, formada pelo ajuste de um modelo de regressão entre os resíduos quadráticos (ou as variâncias amostrais) e os fatores do processo, a qual permite modelar os efeitos dos ruídos não incluídos no experimento de laboratório. O ajuste da segunda equação pode ser feito pelo método S2 (quando houver replicações) ou pelo método RQ, tal como discutido nas seções anteriores.

Wolfinger e Tobias (1998) apresentam uma extensão do modelo (2.38), incluindo também fatores aleatórios propriamente ditos. Estes fatores são aqueles cujos níveis podem ser considerados como sendo aleatoriamente selecionados de uma

população infinita. São fatores tais como canteiros em experimentos agrícolas ou as fornadas e lotes em experimentos industriais, que entram nos experimentos devido a limitações para se considerarem, em condições idênticas, todas as combinações de níveis desejadas. Wolfinger e Tobias descrevem a função de verossimilhança deste modelo ampliado e um método para maximizá-la, permitindo-se, assim, encontrar simultaneamente todas as estimativas de interesse.

Várias aplicações têm sido realizadas com modelos que incluem os ruídos como variáveis independentes na equação de regressão da resposta. Por exemplo, Welch et al. (1990) e Welch et al. (1992) apresentam aplicações em experimentos simulados por computador, Schoemaker et al. (1991) ilustram o método em um experimento de circuitos integrados e Steiner e Hamada (1997) aplicam esta metodologia em experimentos com mistura.

2.3- CONSTRUÇÃO DE MODELOS COM MÚLTIPLAS RESPOSTAS

Em grande parte dos produtos e processos, a qualidade não pode ser avaliada por apenas uma característica funcional do produto. Myers e Montgomery (1995, p. 244) argumentam que a qualidade de um remédio pode ser avaliada preliminarmente pela eficácia com que ele combate a doença, porém não se pode deixar de levar em conta os seus efeitos colaterais. A qualidade de um alimento pode ser avaliada pelo seu gosto, mas seus nutrientes e substâncias prejudiciais à saúde também devem fazer parte da avaliação. Desta forma, estes experimentos produzem múltiplas respostas e, geralmente, correlacionadas.

A análise dos dados de um experimento de múltiplas respostas requer considerações cuidadosas sobre a natureza multivariada dos dados. As variáveis respostas não devem ser investigadas individualmente e independentemente. Inter-relações entre as várias respostas podem levar a análise univariada a conclusões sem sentido (Khuri e Cornell, 1996, p. 251).

Algumas vezes, o conhecimento do produto ou processo pode levar à escolha de duas ou três variáveis contínuas que medem adequadamente o desempenho. Mas, em outras vezes, o número de respostas é extremamente grande, podendo incluir características mensuradas a nível ordinal ou nominal. Neste caso, pode-se fazer um

levantamento prévio de dados e, sobre os dados observacionais, aplicar técnicas que permitam reduzir o dimensionamento, como também, compor escalas numéricas a partir de escalas nominais e ordinais. Técnicas desta natureza são abordadas em Gnanadesikan (1977) e Johnson e Wichern (1992), mas não serão exploradas neste trabalho.

A quase totalidade dos trabalhos de otimização experimental de produtos e processos com múltiplas respostas tratam as respostas de forma isolada, na fase de construção dos modelos de regressão. Este processo pode ser ineficiente, especialmente se as respostas forem fortemente correlacionadas. Esta seção mostra como as construções dos modelos podem ser feitas de forma sistêmica, isto é, sob um processo multivariado, onde a construção de cada equação de regressão pode sofrer influência de respostas de outras equações.

Box et al. (1973) chamam a atenção sobre a possível existência de diferentes tipos de dependência em dados multivariados. Como é usual em análise de regressão, os autores supõem que cada observação pode ser expressa por dois termos: o valor esperado e o termo aleatório ou *erro*. E citam três tipos de dependência: (a) entre os erros, (b) entre os valores esperados das respostas ou (c) nos dados.

O primeiro caso (dependência ou *correlação* entre os erros das respostas) parece ser mais geral. Por exemplo, ao se considerar a porosidade (y_1) e a resistência (y_2) de um material específico, pode-se argumentar que na medida em que o material é mais poroso ele tende a ter menor resistência (correlação negativa), mesmo considerando os fatores controláveis do processo de fabricação fixos. É este tipo de dependência que será tratado nesta seção e no capítulo 4, com a finalidade de promover maior eficiência na estimação dos parâmetros e maior sensibilidade nos testes estatísticos.

Sobre a dependência entre os valores esperados das respostas, Box et al. (1973) citam estudos de mecanismos químicos (estequiometria), onde espera-se, em cada ensaio, algumas relações teóricas entre os valores esperados das componentes da mistura, como por exemplo, que a soma do número de moléculas-grama permaneça constante. Estes relacionamentos entre os valores esperados das respostas induzem dependências também nos dados observados. Este último tipo de dependência não

deve ser funcional devido a presença dos erros aleatórios. Contudo, se a análise for realizada sem se levar em conta tais relações, pode-se chegar a resultados sem sentido prático. Na construção dos modelos de regressão, os autores recomendam eliminar respostas que sejam (aproximadamente) combinações lineares de outras e sugerem um estudo prévio sobre os autovalores e autovetores da matriz de variâncias e covariâncias das respostas, para se identificarem possíveis relações lineares entre os valores esperados das respostas.

2.3.1 - Regressão multivariada

Na literatura de estatística e especialmente na sub-área de análise multivariada, costuma-se chamar de regressão multivariada a modelos com múltiplas variáveis dependentes (respostas) associadas a um conjunto de variáveis independentes (fatores). Em geral, assume-se que todas as equações tenham a mesma forma funcional e variâncias constantes ao longo das observações (homocedasticidade). Desta forma, as equações podem ser representadas por

$$\begin{aligned} y_{1i} &= x_i' \beta_1 + \sigma_1 e_i, \\ y_{2i} &= x_i' \beta_2 + \sigma_2 e_i, \\ &\dots\dots\dots \\ y_{Li} &= x_i' \beta_L + \sigma_L e_i, \end{aligned} \quad i = 1, 2, \dots, n \quad (2.44)$$

onde: $y_{\ell i}$ são os dados oriundos do experimento ou os resultados de transformações aplicadas aos dados;

x_i são os vetores formados pelas linhas da matriz de planejamento X , com dimensão $n \times p$;

β_ℓ é o vetor de parâmetros da ℓ -ésima equação de regressão, com dimensão $p \times 1$;

σ_ℓ^2 são as variâncias das respostas;

e_i são os erros aleatórios, que supostamente têm valores esperado nulos, variâncias unitárias e, preferencialmente, distribuições normais;

Observa-se em (2.44) que a matriz de planejamento é a mesma em todas as equações ($X_1 = X_2 = \dots = X_\ell = X$), caracterizando a mesma forma funcional. Além disso, em cada equação, a variância é constante ao longo das observações ($\sigma_{\epsilon_i}^2 = \sigma_\ell^2$,

$i = 1, 2, \dots, n; \ell = 1, 2, \dots, L$). Nota-se que não se está diferenciando as observações de replicações ou de diferentes pontos experimentais, mas considerando as $n = Mm$ observações da mesma maneira.

O modelo de regressão multivariada (2.44) ainda pode ser apresentado de forma mais compacta, tal como

$$Y = X\beta + \varepsilon \quad (2.45)$$

onde: $Y = (y_1 : y_2 : \dots : y_L)$ é uma matriz $n \times L$, composta das observações das L respostas;

X é a matriz de planejamento com dimensão $n \times p$;

$\beta = (\beta_1 : \beta_2 : \dots : \beta_p)$ é uma matriz $p \times L$, composta pelos vetores de parâmetros e

ε é uma matriz $n \times L$, formada pelos vetores de erros aleatórios, $\sigma_{\ell}\{e_i\}$ ($i = 1, 2, \dots, n; \ell = 1, 2, \dots, L$)

A estimação da matriz de parâmetros β pode ser feita pelos dois métodos mais comuns: mínimos quadrados ou máxima verossimilhança, sendo que no segundo assume-se distribuição normal multivariada. Ambos os métodos produzem o mesmo estimador, dado por (ver Johnson e Wichern, 1992, seção 7.7)

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (2.46)$$

que pode ser decomposto em termos de cada equação, como segue:

$$\hat{\beta}_{\ell} = (X'X)^{-1} X'y_{\ell} \quad \ell = 1, 2, \dots, L. \quad (2.47)$$

Verifica-se que a expressão (2.47) corresponde ao estimador de mínimos quadrados (ou de máxima verossimilhança) de cada equação isoladamente. Portanto, no modelo clássico de regressão multivariada, o processo de estimação de β decompõe-se em problemas univariados, sem influência de possíveis correlações entre as L respostas. Conforme argumenta Khattree e Nalk (1995, p. 69 e 70), o estimador (2.46) é não viciado e, além disso, a variância total de seus componentes é a menor dentre qualquer outro estimador linear e não viciado de β . Ou seja, tem-se uma generalização do Teorema de Gauss - Markov.

A estrutura multivariada é levada em conta na estimação dos erros padrão das estimativas e nos testes estatísticos sobre o modelo. Para tanto, necessita-se estimar a matriz de variâncias e covariâncias Σ do vetor de erros aleatórios. Esta matriz tem dimensão $L \times L$ e a seguinte forma:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1L} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2L} \\ \dots & \dots & \dots & \dots \\ \sigma_{1L} & \sigma_{2L} & \dots & \sigma_{LL} \end{bmatrix} \quad \begin{array}{l} \text{onde } \sigma_{\ell\ell} (\ell = 1, 2, \dots, L) \text{ são as variâncias e } \sigma_{\ell k} = \sigma_{k\ell} \\ (\ell \neq k = 1, 2, \dots, L) \text{ são as covariâncias.} \end{array}$$

A matriz Σ pode ser estimada de forma eficiente e não viciada por

$$S = \frac{1}{n-p} (Y - X\hat{\beta})'(Y - X\hat{\beta}) \quad (2.48)$$

e a decomposição da análise de variância do modelo pode ser feita de forma análoga ao caso univariado, em termos da forma quadrática devida à regressão (R) e da forma quadrática devida ao erro (E), tal como segue (ver detalhes em Khattree e Nalk, 1995, p. 70):¹³

$$T = R + E \quad (2.49)$$

onde: $T = Y'Y - (1/n)Y'1_n1_n'Y$ sendo 1_n o vetor de dimensão n , formado de 1;

$$R = Y'X\hat{\beta} - (1/n)Y'1_n1_n'Y \quad \text{e}$$

$$E = Y'Y - Y'X\hat{\beta}$$

Um teste multivariado do modelo pode ser realizado utilizando-se a chamada estatística *lambda de Wilks*, dada por

$$\Lambda = \frac{|E|}{|T|} \quad (2.50)$$

onde $|E|$ e $|T|$ representam os determinantes das matrizes E e T , respectivamente. Com transformações adequadas de Λ , chega-se a uma distribuição aproximada da qui-quadrado com pL graus de liberdade e, para alguns casos particulares, distribuição exata F (ver detalhes em Johnson e Wichern, 1992, p. 248). Khattree e Nalk (1995, seção 3.5) apresentam a forma linear geral dos testes estatísticos, permitindo fazer, na abordagem multivariada, testes estatísticos similares aos

¹³ Na regressão multivariada está se usando a expressão “forma quadrática” no lugar de “soma quadrática” porque o resultado é uma matriz e não um número.

univariados. Além disso, estes autores orientam como usar o sistema SAS (*statistical analysis systems*) para realizar estes testes. O teste univariado de falta de ajuste também pode ser generalizado para o caso multivariado (ver Khuri e Cornell, 1996, p. 274).

Ressalta-se o aspecto mais genérico da interpretação de um teste multivariado. Nesta abordagem, a hipótese nula aponta para a ausência de efeito de um fator (ou grupo de fatores) sobre o conjunto das respostas como um todo. Assim, a hipótese nula é rejeitada quando houver evidência do efeito sobre alguma das respostas.

Khattree e Nalk (1995, p. 85) ilustram a aplicação da regressão multivariada em um problema de RSM com múltiplas respostas. Os dados provêm de um experimento central composto de segunda ordem, com de três fatores e quatro respostas. Para selecionar a forma funcional do sistema, os autores realizam três testes multivariados, correspondentes às seguintes hipóteses nulas:

$H_0^{(1)}$: o modelo multivariado contém apenas termos lineares (incluindo o vetor de constantes);

$H_0^{(2)}$: o modelo multivariado é quadrático, mas sem termos de interação;

$H_0^{(3)}$: o modelo multivariado contém termos lineares (incluindo o vetor de constantes) e termos de interação de segunda ordem;

A análise dos dados levou à rejeição das hipóteses $H_0^{(1)}$ e $H_0^{(2)}$ e à aceitação de $H_0^{(3)}$, o que indica um sistema de equações que contém termos lineares e quadráticos puros, mas não contém interações entre os fatores.

Observa-se que a realização de testes multivariados acarreta num modelo final onde todas as equações do sistema têm a mesma forma funcional, constituídas do mesmo conjunto de efeitos. Se, ao invés disso, fossem aplicados testes em cada equação isoladamente, poderia resultar em equações com diferentes formas funcionais. A literatura sobre o assunto sugere que, quando puder ser admitida a suposição de que todas as equações têm a mesma forma funcional, os testes multivariados são mais adequados do que os seus similares univariados.

2.3.2 - Sistemas de equações de regressão

Zellner (1962) apresenta uma forma de se tratar com sistemas de equações quando estas são relacionadas apenas pelas correlações dos termos de erro. Este tipo

de sistema é conhecido como SUR (*seemingly unrelated regressions*), que diferencia-se do modelo estudado na seção anterior por permitir diferentes formas funcionais para as equações. Neste contexto, o sistema SUR pode ser considerado como uma generalização da regressão multivariada (ver Press, 1972, p. 219). A ênfase no sistema SUR é no processo de estimação dos parâmetros. Sob certas condições, esta abordagem obtém estimadores mais eficientes do que os obtidos com o tratamento de cada equação isoladamente.

Sejam as L equações do sistema:

$$y_{\ell i} = x_{\ell i}' \beta_{\ell} + \sigma_{\ell} e_{\ell i} \quad i = 1, 2, \dots, n; \ell = 1, 2, \dots, L \quad (2.51)$$

Estas equações podem ser colocadas numa única equação ampliada, como segue:

$$y = X \beta + \varepsilon \quad (2.52)$$

onde:

$y = (y_{11}, y_{12}, \dots, y_{1n}, y_{21}, y_{22}, \dots, y_{2n}, \dots, y_{L1}, y_{L2}, \dots, y_{Ln})'$ é o vetor composto pelas nL observações, agregadas sequencialmente;

$X = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & X_L \end{bmatrix}$ é a matriz de planeamento ampliada, de dimensão nL x P (sendo $P = \sum_{\ell=1}^L p_{\ell}$), formada pelas matrizes de planeamento X_{ℓ} , com dimensões n x p_{ℓ} ($\ell = 1, 2, \dots, L$) e matrizes nulas, 0, de dimensões compatíveis;

$\beta = (\beta_1', \beta_2', \dots, \beta_L')' = (\beta_{11}, \dots, \beta_{1p_1}, \beta_{21}, \dots, \beta_{2p_2}, \beta_{L1}, \dots, \beta_{Lp_L})'$ é o vetor de parâmetros com P elementos, isto é, os parâmetros das L equações agregados sequencialmente num único vetor e

$\varepsilon = (\varepsilon_{11}, \dots, \varepsilon_{1n}, \varepsilon_{21}, \dots, \varepsilon_{2n}, \dots, \varepsilon_{L1}, \dots, \varepsilon_{Ln})' = (\sigma_1 e_{11}, \dots, \sigma_1 e_{1n}, \sigma_2 e_{21}, \dots, \sigma_2 e_{2n}, \dots, \sigma_L e_{L1}, \dots, \sigma_L e_{Ln})'$ é o vetor de erros aleatórios, também ampliado.

Nota-se que, ao efetuar a operação matricial, tem-se exatamente as L equações apresentadas em (2.51). Contudo, a colocação do sistema de equações na forma de vetores e matrizes ampliados não é apenas uma questão de apresentação. A

idéia de Zellner é aplicar o método dos mínimos quadrados generalizados (MQG) sobre a equação matricial ampliada (2.52), permitindo que as correlações entre os erros das diferentes variáveis dependentes influenciem o processo de estimação.

Para se entender o MQG sobre o sistema (2.52), considere apenas duas equações ($L = 2$) e o primeiro par de observações (y_{11}, y_{21}), que carregam os erros aleatórios ($\varepsilon_{11}, \varepsilon_{21}$), com matriz de variâncias e covariâncias $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$. A

figura 2.5 ilustra possíveis formas das densidades de ($\varepsilon_{11}, \varepsilon_{21}$).

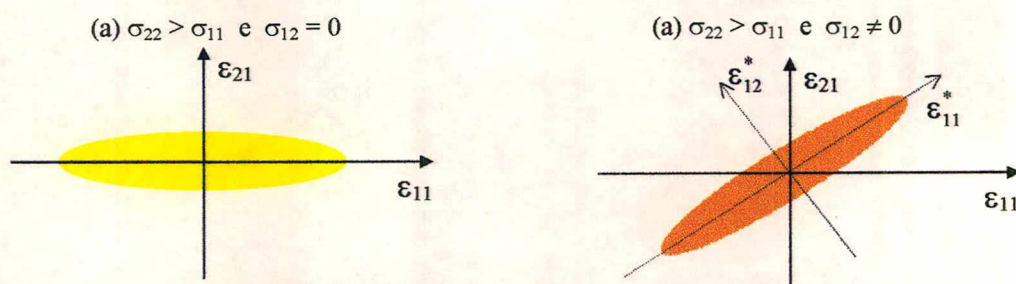


Fig. 2.5 - Ilustração da densidade de probabilidade dos termos de erro de um par de medidas.

O princípio dos mínimos quadrados ordinários baseia-se na soma de quadrados dos erros, onde a parte correspondente às duas medidas consideradas é dada por

$$Sq_1 = \varepsilon_{11}^2 + \varepsilon_{21}^2 = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{21} \end{bmatrix} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \end{bmatrix} \quad (2.53)$$

No presente contexto, as variâncias dos dois termos de erro não devem ser iguais, pois são medidas de características diferentes e, provavelmente, com unidades diferentes. E, como pode-se observar na figura 2.5(a), o termo com maior variância tende a ter peso maior em Sq_1 . A correção na soma de quadrados, em termos das diferentes variâncias, pode ser feita como segue:¹⁴

$$Sq_1^* = \frac{\varepsilon_{11}^2}{\sigma_{11}} + \frac{\varepsilon_{21}^2}{\sigma_{22}} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{21} \end{bmatrix} \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \end{bmatrix} \quad (2.54)$$

¹⁴ Este é o procedimento usual do MQG para a correção da variância não constante nos termos de erro (heterocedasticidade), conforme abordado na seção 2.2.4.

Quando existe correlação entre as medidas ($\sigma_{12} \neq 0$), conforme ilustra a figura 2.5(b), uma soma do tipo (2.54) torna-se mais adequada em termos dos eixos rotacionados (ε_1^* , ε_2^*). Isto pode ser alcançado analiticamente ao se utilizar a seguinte métrica:

$$Sq_1^{**} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{21} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \end{bmatrix} \quad (2.55)$$

No sistema SUR, utiliza-se a soma de quadrados com a métrica apresentada na expressão (2.55). Mas, para contemplar as n observações de L variáveis, onde $\sigma_{\ell k} \neq 0$ ($\ell \neq k = 1, 2, \dots, L$) para as observações da mesma unidade experimental e $\sigma_{\ell k} = 0$ para as observações de unidades experimentais diferentes, a matriz, de dimensão $nL \times nL$, que corrige a métrica da soma de quadrados, deve ter a seguinte forma:

$$\Sigma^* = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 & \dots & \sigma_{1L} & 0 & \dots & 0 \\ 0 & \sigma_{11} & \dots & 0 & \dots & 0 & \sigma_{1L} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{11} & \dots & 0 & 0 & \dots & \sigma_{1L} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{1L} & 0 & \dots & 0 & \dots & \sigma_{LL} & 0 & \dots & 0 \\ 0 & \sigma_{1L} & \dots & 0 & \dots & 0 & \sigma_{LL} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{1L} & \dots & 0 & 0 & \dots & \sigma_{LL} \end{bmatrix} \quad (2.56)$$

Usando o operador \otimes (produto de Kroneker ou produto direto), a matriz Σ^* pode ser representada por

$$\Sigma^* = \Sigma \otimes I_n \quad (2.57)$$

onde I_n é a matriz identidade de ordem n . O produto de Kroneker implica que cada elemento da matriz Σ , de dimensão $L \times L$, seja multiplicado pela matriz I_n , resultando na matriz Σ^* , de dimensão $nL \times nL$.

Admitindo-se Σ^* conhecida e aplicando-se MQG sobre o modelo ampliado (2.52), tem-se o seguinte estimador de β (ver, por exemplo, Judge et al., 1988, p. 250):

$$\tilde{\beta} = [X'(\Sigma^*)^{-1}X]^{-1} X'(\Sigma^*)^{-1}y \quad (2.58)$$

Em termos computacionais, é mais fácil inverter primeiro Σ e, depois, construir $(\Sigma^*)^{-1}$, pois o produto de Kroneker tem a seguinte propriedade: $(\Sigma \otimes I_n)^{-1} = \Sigma^{-1} \otimes I_n$ (ver, por exemplo, Press, 1972, p. 33). O estimador $\tilde{\beta}$ pode ser escrito, em função da matriz Σ , da seguinte maneira:

$$\tilde{\beta} = [X'(\Sigma^{-1} \otimes I_n)X]^{-1} X'(\Sigma^{-1} \otimes I_n)y \quad (2.59)$$

Existem duas situações em que os vetores que compõem $\tilde{\beta}$ coincidem com os estimadores $\hat{\beta}_\ell$ ($\ell = 1, 2, \dots, L$), obtidos por MQO, aplicado a cada equação isoladamente:

- a) quando as correlações entre os termos de erro das variáveis dependentes são nulas, isto é, $\text{Corr}\{\varepsilon_\ell, \varepsilon_k\} = \rho_{\ell k} = 0, \forall \ell, k = 1, 2, \dots, L$ ou
- b) quando o conjunto de variáveis independentes é o mesmo para todas as equações, isto é, $X_\ell = X_k, \forall \ell, k = 1, 2, \dots, L$.

A situação (a) parece bastante natural, pois se não existe correlação entre as respostas, não há razão para que uma resposta influencie na estimação dos parâmetros da equação de outra resposta. Quanto à situação (b), se as matrizes de planejamento das diferentes equações forem iguais, pode-se adotar o modelo de regressão multivariada, onde o processo multivariado de estimação dos coeficientes decompõe-se em processos univariados (ver seção anterior). Uma demonstração formal pode ser vista em Judge et al. (1988, p. 250).

Na prática, a matriz Σ^* não é conhecida e necessita ser estimada. Para se ter uma estimativa desta matriz, pode-se aplicar MQO sobre cada equação isoladamente, obtendo-se os vetores de resíduos brutos $\hat{\varepsilon}_\ell$, associados a cada equação ($\ell = 1, 2, \dots, L$). As covariâncias $\sigma_{k\ell}$ ($\ell, k = 1, 2, \dots, L$) podem ser estimadas por

$$s_{k\ell} = \frac{\hat{\varepsilon}_\ell' \hat{\varepsilon}_k}{\sqrt{(n - p_\ell)(n - p_k)}} \quad (2.60)$$

onde p_ℓ e p_k são os números de parâmetros das equações de regressão y_ℓ e y_k , respectivamente. Outras alternativas para a estimação de $\sigma_{k\ell}$ podem ser encontradas

em Shrivastava e Giles (1987). Chamando de S a matriz formada pelas estimativas das variâncias e covariâncias, tem-se o estimador do vetor de parâmetros do sistema

$$\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2', \dots, \hat{\beta}_L')' = [X'(S^{-1} \otimes I_n)X]^{-1} X'(S^{-1} \otimes I_n)y \quad (2.61)$$

Com a estimativa dos coeficientes, pode-se calcular novos resíduos, estimar novamente a matriz de variâncias e covariâncias e repetir o processo iterativamente. Magnus (1978) mostra que este processo converge para o estimador de máxima verossimilhança de β , desde que os erros tenham distribuição normal multivariada.

Admitindo-se que a distribuição dos erros seja simétrica e que a média exista, o estimador $\hat{\beta}$ de Zellner é não viciado (Kakwani, 1967). Srivastava e Maekawa (1995) mostram que, independentemente da distribuição dos erros, sempre que as equações possuírem o termo constante, o estimador $\hat{\beta}$ de Zellner é aproximadamente não viciado. Além disso, a distribuição de qualquer elemento de $\hat{\beta}$ converge para uma distribuição normal. Estas propriedades independem do número de iterações do processo de estimação.

Quanto à eficiência, se as correlações forem não nulas e se $X_\ell \neq X_k$ para algum $\ell, k = 1, 2, \dots, L$, os estimadores $\hat{\beta}_\ell$ ($\ell = 1, 2, \dots, L$) de Zellner são assintoticamente mais eficientes do que os estimadores produzidos pelo método usual de MQO aplicado a cada equação isoladamente (ver Srivastava e Giles, 1987 e Judge et al., 1988, p. 452). Com simulações do tipo Monte Carlo, Mehta e Swamy (1976) mostram que, num sistema de duas equações, os estimadores de Zellner são notadamente mais eficientes do que os de MQO nas situações em que as correlações são moderadas ou fortes e as equações têm formas funcionais razoavelmente diferentes. Além disso, a eficiência do método SUR é maior quando se têm amostras grandes.

A vantagem do método SUR de Zellner, em relação ao método MQO univariado, aumenta na medida em que se tem correlações fortes entre as variáveis dependentes e equações bem diferentes em relação ao conjunto de variáveis independentes. Por outro lado, se as correlações forem muito baixas ou se $X_\ell = Z_k$, $\forall \ell, k = 1, 2, \dots, L$, o método de Zellner tende a produzir estimativas piores do que o MQO (ver comentários de Srivastava e Giles, 1987, p. 125).

A aplicação de MQG iterativo no sistema SUR tem algumas semelhanças com o MQG iterativo nos modelos de variância (seção 2.2.4). Srivastava e Giles (1987, p. 163) ponderam que, às vezes, o processo iterativo pode levar a perda de eficiência. Situações favoráveis aos estimadores iterativos ocorrem quando houver correlações fortes entre as respostas e equações com termos bem diferentes.

Com a estimativa $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2', \dots, \hat{\beta}_L')'$ do sistema SUR, pode-se calcular os resíduos do sistema e, a partir deles, construir intervalos de confiança e testes de hipóteses para os parâmetros. O teste estatístico univariado, baseado na hipótese linear geral, conforme descrito na seção 2.1.2, pode ser generalizado para o sistema SUR. Sejam $H_0: R\beta = r$ e $H_1: R\beta \neq r$,

onde: β é o vetor ampliado de dimensão $pL \times 1$, composto pelos de parâmetros do modelo SUR;

R é uma matriz de dimensão $J \times pL$, que especifica as J relações de interesse entre os parâmetros e

r é um vetor de dimensão $J \times 1$, composto de valores numéricos, os quais completam as relações entre os parâmetros.

A hipótese H_0 pode ser testada pela estatística:

$$\hat{F}_0 = \frac{(R\hat{\beta} - r)' \left\{ R[X'(S^{-1} \otimes I_n)X]^{-1} R' \right\}^{-1} (R\hat{\beta} - r)}{J} \quad (2.62)$$

Sob a hipótese dos dados serem normalmente distribuídos, \hat{F}_0 tem distribuição aproximada F com J e $nL - P$ graus de liberdade. Observa-se que a aproximação é em termos assintóticos (quando n tende para infinito). Logo, o teste pode não ser válido para amostras pequenas. (ver detalhes em Judge et al., 1988, p. 456).

2.4 - OTIMIZAÇÃO E SOLUÇÕES DE COMPROMISSO

Esta seção apresenta alguns métodos que permitem combinar o objetivo de ajustar a média ao alvo e, ao mesmo tempo, minimizar a variabilidade, considerando as equações de regressão da média e da variância. A seção 2.4.1 aborda o problema quando se tem apenas uma característica de qualidade, ou seja, uma resposta. E a seção 2.4.2 generaliza para o caso de múltiplas respostas.

2.4.1 - Uma característica de qualidade

Muitas vezes, os fatores de entrada são divididos em dois grupos, $x = (x_d, x_a)$, onde o vetor x_d é formado por fatores que alteram a dispersão e o vetor x_a é formado por fatores que só alteram a média do processo. Isto permite um procedimento de otimização em dois estágios (ver León et al., 1987):

- a) busca-se $x_d = x_d^*$ que minimiza a variabilidade e
- b) ajusta-se a média do processo ao alvo em termos dos fatores de x_a , com x_d fixado em x_d^* .

No presente contexto, os fatores do vetor x_a são também chamados de *fatores de ajuste*. Este procedimento em dois estágios tem a vantagem de permitir que certas mudanças no produto ou processo possam ser feitas somente com alterações em x_a , pois, mantendo-se x_d em x_d^* , garante-se variabilidade mínima.

Vining e Myers (1990) apresentam uma metodologia alternativa para ajustar a média ao alvo e minimizar a variância, utilizando-se da abordagem da superfície de resposta dual de Myers e Carter (1973). Esta abordagem consiste em encontrar, sob a região experimental, o ótimo de uma equação primária, condicionada a certas restrições de uma equação secundária. Mais especificamente, sejam $\hat{\mu}(x) = f(x, \hat{\beta})$ e $\hat{\sigma}^2(x) = h(x, \hat{\theta})$ as equações de regressão estimadas da média e da variância, respectivamente. Vining e Myers (1990) propõem usá-las como equações primária e secundária, dependendo da forma da característica de qualidade, tal como segue:

- a) se a característica for do tipo em que existe um valor ideal (*nominal-é-melhor*), então condiciona-se a equação da média a este valor ideal e minimiza-se a variância;
- b) se a característica for do tipo em que quanto maior indica melhor desempenho (*maior-é-melhor*), então condiciona-se a equação da variância a um nível considerado bom e maximiza-se a média;
- c) se a característica for do tipo em que quanto menor indica melhor desempenho (*menor-é-melhor*), então condiciona-se a equação da variância a um nível adequado e minimiza-se a média;

Em termos matemáticos, as três situações podem ser descritas da seguinte maneira:

$$\text{a) } \min_x (\hat{\sigma}^2) = \min_x [h(x, \hat{\theta})] \quad \text{sujeito a} \quad \hat{\mu} = f(x, \hat{\beta}) = \mu_0 \quad \text{e} \quad x \in D$$

$$\text{b) } \max_x (\hat{\mu}) = \max_x [f(x, \hat{\beta})] \quad \text{sujeito a} \quad \hat{\sigma}^2 = h(x, \hat{\theta}) \leq \sigma_0^2 \quad \text{e} \quad x \in D$$

$$\text{c) } \min_x (\hat{\mu}) = \min_x [f(x, \hat{\beta})] \quad \text{sujeito a} \quad \hat{\sigma}^2 = h(x, \hat{\theta}) \leq \sigma_0^2 \quad \text{e} \quad x \in D$$

onde: μ_0 é um valor fixo (o valor nominal desejável, por exemplo);

σ_0^2 é um valor fixo (a variabilidade máxima aceitável como boa) e

D é a região experimental.

Nos casos (b) e (c), pode-se usar diferentes valores de σ_0^2 para estudar o comportamento do máximo (ou mínimo) na expressão da média. O mesmo pode ser feito com μ_0 , no caso (a).

Myers e Montgomery (1995, p. 502) observam que o método não só permite chegar a uma solução de compromisso entre ajustar a média ao alvo e reduzir a variabilidade, como também oferecer ao pesquisador uma grande aprendizagem sobre o processo, pois é avaliado o nível ótimo de operação relativo à variável principal, para diferentes níveis da variável secundária. Isto fica mais evidente quando se tem apenas dois ou três fatores significativos, o que permite colocar num mesmo gráfico as curvas de nível das duas equações.

Castilho e Montgomery (1993), Copeland e Nelson (1996) e Castilho et al. (1997) apresentam técnicas de programação não linear para a solução do problema de Vining e Myers. Copeland e Nelson (1996) também sugerem um certo relaxamento nas restrições, considerando o fato de se trabalhar com estimativas das superfícies de resposta e não com as funções verdadeiras, que são desconhecidas. Por exemplo, para o caso (a), eles sugerem

$$\min_x (\hat{\sigma}^2) = \min_x [h(x, \hat{\theta})] \quad \text{sujeito a} \quad (\hat{\mu} - \mu_0)^2 \leq \Delta^2 \quad \text{e} \quad x \in D$$

onde Δ é metade da amplitude do intervalo de confiança de $\hat{\mu}$.

Um método usado por muitos autores (ver, por exemplo, Lin e Tu, 1995 e Steiner e Hamada, 1997) é a minimização direta do erro quadrático médio (ou valor esperado da função perda quadrática), sob a região experimental. Seja τ o valor desejado para a resposta y , a qual depende dos fatores de entrada x , ou seja: $y = y(x)$. O erro quadrático médio é dado por

$$EQM(x) = [y(x) - \tau]^2 = [\mu(x) - \tau]^2 + \sigma^2(x) \quad (2.63)$$

sendo que a média e a variância do processo, $\mu(x)$ e $\sigma^2(x)$, são substituídas por suas respectivas estimativas, ou seja:

$$\hat{\mu}(x) = f(x, \hat{\beta}) \quad e \quad \hat{\sigma}^2(x) = h(x, \hat{\theta}) \quad (2.64)$$

Ribeiro e Elsayed (1995) consideram uma função objetivo que agrega os três seguintes objetivos, considerados típicos na engenharia da qualidade:

- a) minimizar desvios do alvo;
- b) maximizar a robustez em relação aos ruídos e
- c) maximizar a robustez em relação a possíveis flutuações dos parâmetros do processo (fatores controláveis de entrada).

A minimização do EQM, expressão (2.60), atinge aos objetivos (a) e (b). O objetivo (c) também seria atingido se o experimento, executado em laboratório, mantivesse a flutuação dos parâmetros do processo tal qual ocorre no chão de fábrica, o que não é usual na experimentação. Para contornar este problema, Ribeiro e Elsayed (1995) propõem a seguinte função objetivo de três partes, que aqui é adaptada para o caso de resposta única:

$$Z(x) = [\hat{\mu}(x) - \tau]^2 + \hat{\sigma}^2(x) + \sum_{j=1}^k \sigma_{x_j}^2 \left(\frac{\partial \hat{\mu}}{\partial x_j} \right)^2 \quad (2.65)$$

onde $\sigma_{x_j}^2$ é uma avaliação da variância do parâmetro do processo x_j no chão de fábrica ($j = 1, 2, \dots, k$). Observa-se que, para os parâmetros que são mantidos rigorosamente fixos no chão de fábrica, tem-se $\sigma_{x_j}^2 = 0$. Se todos os parâmetros do processo são fixos, a última parte da expressão (2.65) é anulada e tem-se o tradicional EQM, como em (2.62).

Ribeiro e Elsayed justificam a expressão (2.65) através da expansão em série de Taylor de $\hat{\mu}(x)$, considerando a aproximação de primeira ordem e supondo independência entre as flutuações dos parâmetros do processo. Observando o modelo que agrega efeitos de locação e de dispersão, como foi visto na seção 2.2.6, verifica-se que as duas últimas partes da expressão (2.65) são semelhantes à equação do processo da variância, expressão (2.40), quando se consideram os parâmetros do processo que flutuam no chão de fábrica como ruídos. Isto vem a fornecer uma nova justificativa para a expressão (2.65).

Em geral, o estudo não se limita a encontrar a combinação ótima de níveis de operação, mas também conhecer o processo e a sensibilidade dos fatores em torno do ótimo. Considerando que $\hat{\mu}(x) = f(x, \hat{\beta})$ e $\hat{\sigma}^2(x) = h(x, \hat{\theta})$ sejam funções polinomiais ou exponenciais, a análise de EQM(x), ou de Z(x), pode ser feita com as técnicas usuais de análise de uma superfície de resposta (ver seção 2.1.3), pois EQM(x) e Z(x) são deriváveis em toda a região experimental.

2.4.2 - Múltiplas características de qualidade

Nesta seção, serão discutidas algumas técnicas para se encontrar uma solução de compromisso quando se tem múltiplas respostas e, portanto, um sistema de equações para as médias do processo e um sistema de equações para as variâncias do processo.

No caso de apenas uma resposta, os métodos normalmente baseiam-se na idéia de minimização do valor esperado da função perda quadrática (ou, equivalentemente, minimização do erro quadrático médio). No caso de múltiplas respostas, pode-se generalizar a função perda quadrática. Seguindo Kapur e Cho (1996), sejam:

- y um vetor composto de L características de qualidade (respostas) do produto ou processo;
- y_ℓ o ℓ -ésimo elemento do vetor y;
- τ_ℓ o valor especificado (alvo) para a característica y_ℓ ($\ell = 1, 2, \dots, L$);
- k_ℓ o custo, em termos financeiros, do afastamento de y_ℓ de τ_ℓ ($\ell = 1, 2, \dots, L$);

$k_{i\ell}$ o custo, em termos financeiros, dos afastamentos simultâneos entre y_i e τ_i e de y_ℓ e τ_ℓ ($i, \ell = 1, 2, \dots, L$);

A função perda multivariada, baseada na aproximação de segunda ordem da série polinomial de Taylor, é dada por

$$L(y) = \sum_{i=1}^L \sum_{\ell=1}^i k_{i\ell} (y_i - \tau_i)(y_\ell - \tau_\ell) \quad (2.66)$$

E o seu valor esperado por

$$\mu_L = \sum_{\ell=1}^L k_\ell \left[(\mu_\ell - \tau_\ell)^2 + \sigma_\ell^2 \right] + \sum_{i=2}^L \sum_{\ell=1}^{i-1} k_{i\ell} \left[\sigma_{i\ell} + (\mu_i - \tau_i)(\mu_\ell - \tau_\ell) \right] \quad (2.67)$$

onde: μ_ℓ é o valor esperado de y_ℓ ($\ell = 1, 2, \dots, L$);
 σ_ℓ^2 é a variância de y_ℓ ($\ell = 1, 2, \dots, L$) e
 $\sigma_{i\ell}$ é a covariância entre y_i e y_ℓ ($i > \ell = 1, 2, \dots, L$)

Nota-se que, em termos práticos, existem dificuldades para se trabalhar com μ_L , pois necessita-se da avaliação dos custos $k_{i\ell}$ e de estimativas das covariâncias $\sigma_{i\ell}$. Para contornar estas dificuldades, a maioria das funções objetivo baseadas na função perda utilizam somente a primeira parte da expressão (2.64), ou seja, consideram

$$\mu_L = \sum_{\ell=1}^L k_\ell \left[(\mu_\ell - \tau_\ell)^2 + \sigma_\ell^2 \right] \quad (2.68)$$

onde implicitamente está se admitindo:

- a) que existe independência entre as características de qualidade e
- b) que os custos de afastamentos do alvo possam ser avaliados de forma isolada.

Desta forma, para o propósito de minimização, pode-se substituir os custos por pesos relativos de cada y_ℓ ($\ell = 1, 2, \dots, L$). Os pesos relativos também devem levar em conta as diferentes amplitudes das respostas.

A função objetivo proposta por Ribeiro e Elsayed (1995), que na seção anterior foi apresentada para o caso univariado, toma a seguinte forma quando se tem múltiplas características de qualidade:

$$Z(x) = \sum_{\ell=1}^L W_{\ell} \left\{ \left[\hat{\mu}_{\ell}(x) - \tau_{\ell} \right]^2 + \hat{\sigma}_{\ell}^2(x) + \sum_{j=1}^k \sigma_{x_j}^2 \left(\frac{\partial \hat{\mu}_{\ell}}{\partial x_j} \right)^2 \right\} \quad (2.69)$$

onde: W_{ℓ} é a importância relativa da característica de qualidade ℓ ($\ell = 1, 2, \dots, L$);
 $\hat{\mu}_{\ell}(x)$ é a equação de regressão da média da característica de qualidade ℓ ;
 $\hat{\sigma}_{\ell}^2(x)$ é a equação de regressão da variância da característica de qualidade ℓ e
 $\sigma_{x_j}^2$ é uma avaliação da variância do parâmetro do processo x_j no chão de fábrica ($j = 1, 2, \dots, k$).

Ribeiro e Elsayed (1995), Caten (1995) e Pasa et al. (1996) apresentam estudos de caso com múltiplas características de qualidade, usando a função objetivo (2.69). Caten (1995) enfatiza a importância de se analisarem as superfícies de resposta da média e da variância das principais características de qualidade, como forma de aumentar o conhecimento sobre o processo global.

Outra forma de agregar várias respostas e que tem aparecido bastante na literatura é através da chamada função de utilidade (*desirability function*). Este método consiste em transformar as respostas em características adimensionais e, depois, agregá-las através da média geométrica. Conforme o desenvolvimento de Derringer e Suich (1980) e Myers e Montgomery (1995, p. 247), a forma da transformação varia de acordo com a pretensão que se tem a respeito de cada resposta (ser minimizada, ser maximizada ou ser levada a algum valor ideal). Mais especificamente, para as respostas que devem ser maximizadas, a transformação é do tipo [ver figura 2.6 (a)]:

$$d = \begin{cases} 0 & \text{se } \hat{y} < A \\ \left(\frac{\hat{y} - A}{B - A} \right)^s & \text{se } A \leq \hat{y} < B \\ 1 & \text{se } \hat{y} \geq B \end{cases} \quad (2.70)$$

onde: A é o menor valor aceitável para a resposta y ;

B é um valor extremamente bom para y e

s é uma constante positiva que fornece a forma desejável da função.

Para as respostas que devem ser minimizadas, a transformação é do tipo [ver figura 2.6 (b)]:

$$d = \begin{cases} 1 & \text{se } \hat{y} < B \\ \left(\frac{C - \hat{y}}{C - B}\right)^t & \text{se } B \leq \hat{y} < C \\ 0 & \text{se } \hat{y} \geq C \end{cases} \quad (2.71)$$

onde: C é o maior valor aceitável para a resposta y;

B é um valor extremamente bom para y e

t é uma constante positiva que fornece a forma desejável da função.

E, para as respostas que devem ser levadas a um valor nominal ideal, a transformação pode ser feita como [ver figura 2.6 (c)]

$$d = \begin{cases} \left(\frac{\hat{y} - A}{B - A}\right)^s & \text{se } A \leq \hat{y} < B \\ \left(\frac{C - \hat{y}}{C - B}\right)^t & \text{se } B \leq \hat{y} < C \\ 0 & \text{se } \hat{y} < A \text{ ou } \hat{y} \geq C \end{cases} \quad (2.72)$$

onde: A é o menor valor aceitável para a resposta y;

B é o valor ideal para y,

C é o maior valor aceitável para y e

s e t são constantes positivas.

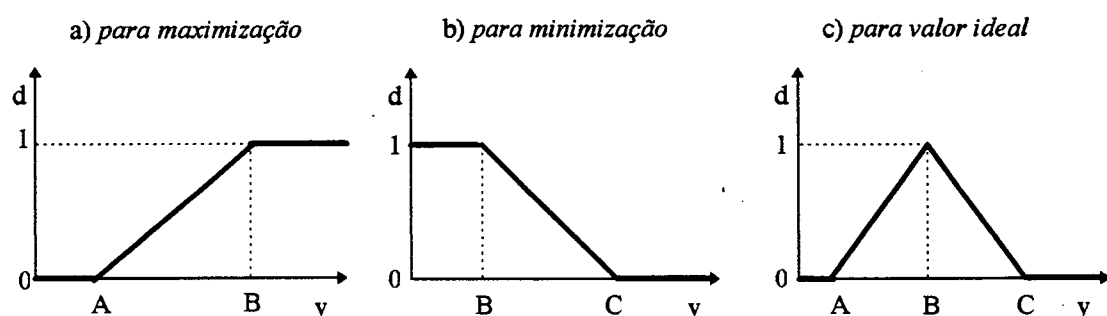


Fig. 2.6 - Diferentes transformadas para a construção da função utilidade, considerando $s = t = 1$.

Como \hat{y} é função dos fatores de entrada, $x = (x_1, x_2, \dots, x_k)'$, isto é, $\hat{y} = \hat{y}(x)$, então d também é função de x, ou seja, $d = d(x)$. Considerando que no problema existam L respostas transformadas, $d_j = d_j(x)$ ($j = 1, 2, \dots, L$), Derringer e Suich (1980) sugerem agregá-las através da média geométrica, ou seja:

$$D(x) = \sqrt[L]{d_1(x)d_2(x) \dots d_L(x)} \quad (2.73)$$

A justificativa para o uso da média geométrica é que pontos x_0 que levam algum y_j ($j = 1, 2, \dots, L$) para valores inaceitáveis acarretam numa média geométrica nula, isto é, $D(x_0) = 0$. Desta forma, evita-se que se tenha solução ótima com alguma resposta em nível inaceitável.

Cabe observar que, onde se usa $\hat{y} = \hat{y}(x)$, como equação de regressão de uma resposta, pode-se também considerar, em seu lugar, as equações da média, $\hat{\mu} = \hat{\mu}(x)$, e da variância, $\hat{\sigma}^2 = \hat{\sigma}^2(x)$, relativas à resposta y . Ou seja, a técnica de Derringer e Suich pode ser naturalmente generalizada para problemas de projeto robusto.

Tendo a função de utilidade $D = D(x)$, pode-se tratar o problema multivariado (de múltiplas respostas) como se fosse univariado (de uma resposta, D). Castilho et al. (1996) apresentam uma pequena alteração na função utilidade, fazendo com que ela seja derivável em todos os pontos. Isto permite utilizar o método de otimização baseado no gradiente, que é usual na metodologia de superfície de resposta (ver seção 2.1.3).

A função perda multivariada e a função utilidade também podem ser aplicadas diretamente sobre as observações do experimento, isto é, para cada vetor de observações, $y_i = (y_{1i}, y_{2i}, \dots, y_{Li})'$, $i = 1, 2, \dots, n$, pode-se calcular o valor da função perda (ou da função utilidade), resultando num problema de resposta única, já na fase de estimação dos modelos. Contudo, o uso das equações de regressão das várias respostas, como discutido nesta seção, permite conhecer melhor o processo, pois, com a análise da superfície de resposta da função perda (ou da função utilidade), acompanhada de equações de cada característica de qualidade, tanto em termos da média, quanto em termos da variância, oferece aos experimentadores uma visão global e, simultaneamente, uma visão sobre cada característica relevante do processo ou sistema.

MODELOS DE VARIÂNCIA EM EXPERIMENTOS COM POUCAS REPLICAÇÕES

Na seção 2.3, foi discutido o problema de se construírem modelos de regressão para a variância de uma resposta y , caracterizada como uma variável aleatória contínua e com a média e a variância associadas ao vetor de fatores do processo, $x = (x_1, x_2, \dots, x_k)'$, da seguinte forma:

$$\mu_y = f(x, \beta) \quad \text{e} \quad \sigma_y^2 = g(x, \theta) \quad (3.1)$$

onde f e g são funções lineares (ou linearizáveis por transformações) em termos dos vetores de parâmetros β e θ , respectivamente. A equação de μ_y é chamada de equação da média do processo e a equação de σ_y^2 é denominada de equação da variância do processo.

Três formas de se obterem estimativas para o parâmetro θ e, conseqüentemente, se ter um modelo para a variância, foram descritas na seção 2.3. A primeira, chamada de S2, é aplicada em experimentos que comportam replicações. Calcula-se a variância amostral em cada ponto experimental e esta é considerada como variável dependente no modelo de regressão. Usualmente, aplica-se uma transformação logarítmica para garantir as propriedades desejáveis de um modelo de regressão e, assim, permitir o uso do método de estimação dos mínimos quadrados ordinários (MQO).

Na segunda forma, denominada de RQ, utilizam-se os resíduos do modelo da resposta. Ou seja, após ajustar o modelo $\mu_y = f(x, \beta)$ aos dados, calculam-se os resíduos: $\hat{e}_i = y_i - \hat{\mu}_i$, onde $\hat{\mu}_i$ é o valor predito pelo modelo ($i = 1, 2, \dots, n$). Conforme discutido em (2.2.4), recomenda-se efetuar uma correção nestes resíduos devido às perdas dos graus de liberdade no modelo de μ_y . Os resíduos quadráticos, \hat{e}_i^2 ($i = 1, 2, \dots, n$), são estimadores das variâncias, podendo-se, assim, serem usados

como observações da variável dependente do modelo de regressão de σ_y^2 . A estimação do vetor de parâmetros θ pode ser feita pela técnica de máxima verossimilhança ou por mínimos quadrados. Neste trabalho, quando existir replicações, considerar-se-á $\hat{\epsilon}_i^2$ como o resíduo quadrático médio do i -ésimo ponto experimental ($i = 1, \dots, M$) e se dará preferência às técnicas baseadas no princípio dos mínimos quadrados.

Na terceira forma, simula-se, em laboratório, os ruídos que podem afetar o desempenho do produto. Com os dados experimentais, ajusta-se um modelo misto para a resposta, incluindo fatores controláveis, ruídos e interações entre fatores controláveis e ruídos. Nesta abordagem, o modelo estimado da resposta não coincide com o modelo estimado da média, pois dele se extraem tanto o modelo da média quanto o da variância (ver seção 2.2.6).

O primeiro método, S2, além da sua simplicidade, pode ser considerado eficiente quando o número de replicações for grande, conforme detalham Bartlett e Kendall (1946). Aparentemente, após a abordagem de Taguchi, é o método mais usado na engenharia da qualidade. Quanto ao método RQ, apesar dele ser teoricamente mais eficiente do que o método S2, especialmente quando se tem poucas replicações, apresenta a desvantagem de que os resíduos só fornecem informações confiáveis sobre a variância do processo se o modelo da resposta for corretamente especificado. O terceiro procedimento é recomendável quando existem ruídos bem conhecidos e que possam ser simulados em laboratório, mas o modelo da variância pode ser complementado com uma parcela gerada por um dos métodos anteriores (ver Engel e Huele, 1996).

Neste capítulo, apresentam-se três procedimentos, dispostos nas seções 3.2, 3.3 e 3.4, que tendem melhorar a eficiência na estimação dos parâmetros do modelo de σ_y^2 , em experimentos com poucas replicações, ou mesmo em experimentos não replicados. Os procedimentos propostos serão avaliadas por simulações do tipo Monte Carlo, cuja descrição geral é feita na seção seguinte.

3.1 - DESCRIÇÃO GERAL DAS SIMULAÇÕES

Os modelos são simulados em termos de uma resposta y , seguindo uma distribuição normal de média e variância descritos pelos seguintes modelos:

$$\mu_y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \beta_{12} x_1 x_2 + \dots + \beta_{k-1,k} x_{k-1} x_k \quad (3.2)$$

$$\sigma_y^2 = \sigma_0^2 \exp\{\theta_1 x_1 + \dots + \theta_k x_k + \theta_{12} x_1 x_2 + \dots + \theta_{k-1,k} x_{k-1} x_k\} \quad (3.3)$$

Os fatores x_1, x_2, \dots, x_k são sempre codificados da forma usual, por exemplo, num experimento do tipo 2^k a codificação é -1 para o nível inferior e $+1$ para o nível superior do fator, sendo que muitas vezes estes níveis codificados são representados somente pelos sinais $-$ e $+$. Em alguns experimentos podem ser incluídos termos quadráticos.

No modelo da variância, o parâmetro $\sigma_0^2 = \exp\{\theta_0\}$ mede a magnitude geral das variâncias e os parâmetros $\theta_1, \theta_2, \dots$ estão associados com o número de vezes que a variância em um nível do fator é superior à variância no outro nível do fator. Por exemplo, $\theta_1 = 0,347$ implica em que a variância no nível superior de x_1 é 2 vezes a variância no nível inferior de x_1 . Os valores para θ_i ($i = 1, 2, \dots$) de 0,549, 0,693 e 0,896 implicam na alteração da variância por fatores de 3, 4 e 6, respectivamente. Nota-se que um fator de 4 vezes na variância corresponde a um fator de 2 vezes no desvio padrão, o qual é medido na mesma unidade dos dados.

Na maioria dos modelos simulados, considerou-se o fator x_1 como o de maior efeito de dispersão e o seu coeficiente, θ_1 , como o parâmetro focal, isto é, o parâmetro onde se estuda a eficiência da estimação. Neste trabalho, a eficiência é calculada da forma tradicional, ou seja, mede-se a eficiência relativa de dois estimadores ($\hat{\theta}_a$ e $\hat{\theta}_b$), de um certo parâmetro θ , pela razão entre os seus erros quadráticos médios. Mais especificamente:

$$\text{Eficiência relativa de } \hat{\theta}_b \text{ em relação a } \hat{\theta}_a = \frac{E\{(\hat{\theta}_a - \theta)^2\}}{E\{(\hat{\theta}_b - \theta)^2\}} \quad (3.4)$$

onde o operador $E\{\cdot\}$ representa valor esperado. No estudo por simulação, os valores esperados são estimados pelas correspondentes médias aritméticas, calculadas sobre um grande número de amostras simuladas.

Os programas computacionais das simulações foram desenvolvidos no *proc IML (Interactive Matrix Language)* do SAS (*Statistical Analysis System*), versão 6.11. Os termos aleatórios foram gerados pela subrotina RANNOR e, nos casos em que havia interesse em controlar a variação experimental, utilizou-se a mesma sequência de sementes, provocando conjuntos idênticos de amostras aleatórias (ou pseudo-aleatórias). Os principais programas computacionais das simulações são apresentados no apêndice. A apresentação dos resultados, por meios gráficos, foi feita com o apoio do STATISTICA, versão 5.

Nas três seções seguintes, descrevem-se experimentos simulados com grande número de fatores. Contudo, tais experimentos foram precedidos de alguns ensaios prévios (não descritos no trabalho), onde ganhou-se alguns conhecimentos que facilitaram na escolha dos fatores e do planejamento experimental.

3.2 - O USO DE FATORES NÃO SIGNIFICATIVOS COMO PSEUDO-REPLICAÇÕES

Foi visto na seção 2.2.3, que nos casos em que a variância amostral é calculada sobre um número muito pequeno de observações, a eficiência de $\log(S^2)$ é bastante baixa. Nestes casos, a aplicação da técnica dos mínimos quadrados, num modelo de regressão baseado em $\log(S^2)$, produz estimadores pouco precisos. O que se propõe é considerar os fatores que não influenciam a variância da resposta como se fossem replicações, aumentando-se, assim, o número de observações usadas no cálculo das variâncias amostrais (ou dos resíduos quadráticos médios).

A identificação de fatores que não provocam variações em σ_y^2 pode ser feita pelas técnicas descritas no capítulo 2, como por exemplo, o gráfico normal dos efeitos. Observa-se que, na grande maioria das aplicações descritas na literatura, poucos fatores são identificados como tendo efeitos de dispersão. Assim sendo, é quase sempre possível efetuar a agregação. A figura 3.1 ilustra o processo de agregação num experimento do tipo fatorial 2^3 , onde as variâncias amostrais dos dois níveis de x_3 são agregadas.

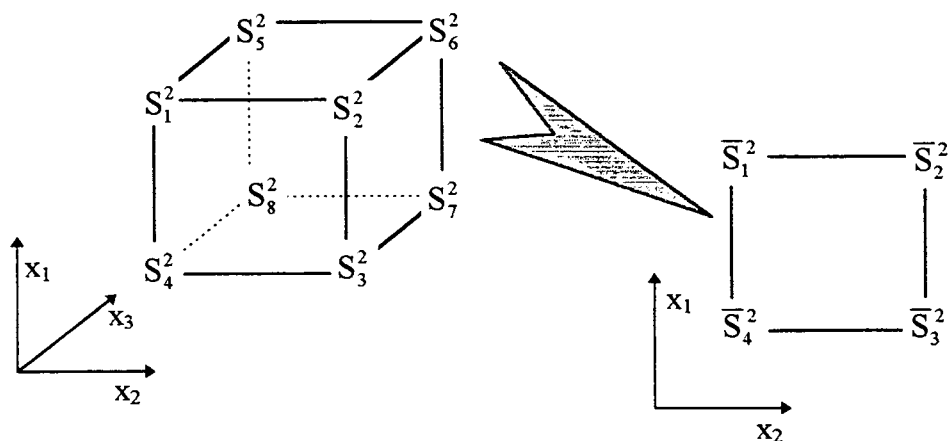


Fig. 3.1 - Ilustração do processo de agregação de variâncias amostrais num experimento fatorial 2^3 .

Quando o número de replicações é idêntico em todos os pontos experimentais, pode-se agregar as variâncias amostrais pela média aritmética. Por exemplo, na ilustração da figura 3.1, \bar{S}_1^2 é a média aritmética de S_1^2 e S_5^2 . De forma mais geral, para agregar as variâncias amostrais S_i^2 ($i = 1, 2, \dots, I$), calculadas com m_i observações, respectivamente, faz-se:

$$\bar{S}^2 = \frac{\sum_{i=1}^I (m_i - 1) S_i^2}{\sum_{i=1}^I (m_i - 1)} \quad (3.5)$$

No método RQ, a agregação também pode ser feita pela média aritmética, mas o mesmo resultado é obtido quando se desconsidera o fator sem efeito no modelo de σ_y^2 e calcula-se os resíduos quadráticos médios nos pontos experimentais formados pelas combinações dos níveis dos outros fatores.

O modelo de regressão é construído em termos de $\log(\bar{S}_\ell^2)$ (método S2) ou $\log(\bar{\epsilon}_\ell^2)$ (método RQ), $\ell = 1, 2, \dots, M'$, onde M' é o número de pontos experimentais formado pelos fatores significativos. Com a agregação tem-se menos pontos para efetuar as estimativas, mas obtém-se melhor avaliação da dispersão nestes pontos. A eficiência deste procedimento será avaliada nas três seções seguintes. Maior ênfase será dada à classe de projetos experimentais fatoriais do tipo 2^k , por ser bastante comum na prática e muito explorada na literatura. Além disso, em projetos fatoriais, o procedimento de agregação não muda a estrutura do projeto, apenas reduz a sua dimensão.

Em experimentos do tipo 2^k , se houver apenas um efeito significativo e a agregação for feita em termos de todos os outros fatores, o processo aqui descrito assemelha-se com o método de Box e Meyer (1986), conforme apresentado na seção 2.2.1. Ressalta-se que o método de Box e Meyer é eficiente na presença de apenas um efeito de dispersão, como aponta Nair e Pregibon (1988). Por outro lado, quando houver mais de um efeito de dispersão, o procedimento proposto pode ser considerado como uma mistura do método de Box e Meyer (1986) com o método clássico de Bartlett e Kendall (1946).

3.2.1 - Experimentos fatoriais 2^k não replicados

Em experimentos não replicados, a estimação dos parâmetros do modelo de σ_y^2 fica restrita ao método RQ. Ao suprimir algum fator, produzindo-se duas replicações artificiais em cada ponto experimental, pode-se tanto usar o método RQ quanto o método S2. Contudo, para se poder comparar a eficiência dos estimadores com o processo de agregação, o estudo restringe-se ao método RQ.

Segue o estudo de simulação nº 1, realizado para se avaliar a eficiência relativa do procedimento proposto. Em todos os casos simulados, o fator suprimido foi o de maior índice (por exemplo, num projeto 2^4 , suprimiu-se x_4). Para evitar confusão com os símbolos, adotaram-se as letras latinas maiúsculas (A, B, C, ...) para representar os fatores do estudo de simulação, enquanto que os fatores do processo continuam sendo representados por x_1, x_2, \dots, x_k . O quadro 3.1 descreve as características deste primeiro estudo de simulação.

Quadro 3.1 - Fatores e níveis usados no estudo de simulação nº 1.

Fator	Descrição	Nível inferior (-)	Nível superior (+)
A	Tipo de experimento / Número de observações	fatorial 2^5 (64 observações)	fatorial 2^4 (16 observações)
B	Número de parâmetros no modelo da média	Modelo gerado e especificado no processo de estimação somente com o termo constante ($\beta_0 = 0$) e efeitos principais (β_1, β_2, \dots), sendo que os valores usados na simulação são descritos em D e H.	Incluem-se, também, interações de segunda ordem, geradas com a metade do valor de β_1 (são 3 termos em 2^4 e 10 termos em 2^5). Não se incluem as interações com o fator a ser suprimido).

continua...

Quadro 3.1 - Fatores e níveis usados no estudo de simulação nº 1 (continuação).

Fator	Descrição	Nível inferior (-)	Nível superior (+)
C	Número de parâmetros no modelo da variância	Modelo gerado e especificado na estimação somente com o termo constante e θ_1 , sendo que o valor de θ_1 , no modelo de simulação, é descrito no fator F.	Incluem-se os outros efeitos principais e interações de segunda ordem, todos gerados com valor 0,347. São 5 termos em 2^4 e 14 termos em 2^6 . Não se incluem as interações com o fator a ser suprimido.
D	Coefficiente do modelo da média, referente ao fator suprimido (β_4 em 2^4 e β_6 em 2^6).	Coefficiente nulo e excluído no processo de estimação.	Coefficiente igual a β_1 e incluído no processo de estimação.
E	Coefficiente do fator suprimido no modelo da variância.	Coefficiente nulo e excluído no processo de estimação.	Coefficiente igual a 0,347, mas não considerado no processo de estimação.
F	Força do efeito de dispersão	$\theta_1 = 0,549$, que altera a variância em 3 vezes	$\theta_1 = 0,896$, que altera a variância em 6 vezes
G	Magnitude geral das variâncias	$\sigma_0^2 = 1$	$\sigma_0^2 = 3$
H	Efeitos de locação	$\beta_0 = 0, \beta_i = 1 (i = 1, 2, \dots)$	$\beta_0 = 8, \beta_1 = 8, \beta_2 = 4, \beta_3 = 8, \beta_4 = 4, \dots$

Observa-se, por exemplo, que quando todos os fatores do quadro 3.1 estão no nível inferior (-1), os dados são simulados segundo as equações:

$$\mu_y = x_1 + x_2 + x_3 + x_4 + x_5 \tag{3.6}$$

$$\sigma_y^2 = \exp\{0,549x_1\} \tag{3.7}$$

conforme um planejamento fatorial 2^6 . No processo de estimação, os modelos são especificados da seguinte forma:

$$\mu_y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 \tag{3.8}$$

$$\sigma_y^2 = \exp\{\theta_0 + \theta_1x_1\} \tag{3.9}$$

e estuda-se a eficiência da estimação do parâmetro θ_1 de duas maneiras: agregando-se e não se agregando os resíduos quadráticos dos dois níveis de x_6 .

O fator A (ver quadro 3.1) não contempla explicitamente experimentos fracionados, porém, em termos de simulação, um experimento fracionado do tipo 2^{6-2} pode ser considerado equivalente a um 2^4 , desde que se mantenha o número de

parâmetros. Assim, os resultados deste estudo podem ser ampliados para experimentos do tipo 2^{k-r} .

Os fatores A, B e C estão associados aos graus de liberdade para se proceder as estimativas, porém sob enfoques diferentes. O fator D foi incluído com o propósito de se verificar o efeito da supressão (na fase de estimação dos parâmetros do modelo de σ_y^2) de um fator influente no modelo da média. O fator E, por sua vez, contempla situações em que não se detecta um efeito real em σ_y^2 e, em consequência, especifica-se o modelo erroneamente. Os fatores F, G e H procuram avaliar até que ponto diferenças nos coeficientes podem afetar o nível de eficiência do procedimento proposto.

Para avaliar o efeito de cada fator do quadro 3.1, realizou-se um projeto experimental com os oito fatores, adotando-se dois fracionamentos (2^{8-2}), o qual leva a observar 64 combinações de níveis dos fatores A, B, ..., H. Este projeto experimental permite estimar todos os efeitos principais e todas as interações de segunda ordem, desde que as interações de ordem superior possam ser negligenciadas. A tabela 3.1 apresenta as 64 combinações dos níveis dos fatores A, B, ..., H, geradas pelo programa *Statistica*, versão 5.0.

Em cada uma das 64 condições experimentais, executaram-se 2.000 realizações simuladas do experimento. As equações de regressão foram estabelecidas em termos do logaritmo dos resíduos quadráticos (ou dos resíduos quadráticos médios, no caso do uso de um fator como *pseudo-replicação*) e o processo de estimação foi feito com 2 iterações (ver detalhes na seção 2.2.4). A partir das estimativas de θ_1 , realizadas com e sem a agregação, calculou-se a eficiência relativa do processo de agregação. Além disso, avaliou-se o vício (diferença média entre o estimador e o parâmetro) no processo com a agregação (ver tabela 3.1). Para maior controle da variância experimental, o conjunto das 2.000 amostras aleatórias simuladas foi o mesmo nos 32 ensaios de cada projeto de experimento analisado.

Tabela 3.1 - Matriz experimental (2^{8-2}) e resultados da eficiência relativa (ef.) e do vício com a agregação (vic.) do estudo de simulação nº 1.

#	A	B	C	D	E	F	G	H	ef.	vic.	#	A	B	C	D	E	F	G	H	ef.	vic.
1	+	+	+	+	+	+	+	+	1,086	-0,068	33	-	+	+	+	+	+	-	-	1,225	-0,013
2	+	+	+	+	+	-	+	-	1,037	-0,011	34	-	+	+	+	+	-	-	+	1,199	0,002
3	+	+	+	+	-	+	+	-	1,090	-0,068	35	-	+	+	+	-	+	-	+	1,210	-0,012
4	+	+	+	+	-	-	+	+	1,042	-0,013	36	-	+	+	+	-	-	-	-	1,202	0,003
5	+	+	+	-	+	+	-	+	0,944	-0,005	37	-	+	+	-	+	+	+	-	1,159	0,009
6	+	+	+	-	+	-	-	-	0,880	-0,001	38	-	+	+	-	+	-	+	+	1,105	0,010
7	+	+	+	-	-	+	-	-	0,957	-0,005	39	-	+	+	-	-	+	+	+	1,186	0,008
8	+	+	+	-	-	-	-	+	0,880	-0,001	40	-	+	+	-	-	-	+	-	1,073	0,007
9	+	+	-	+	+	+	-	+	1,063	-0,052	41	-	+	-	+	+	+	+	-	1,346	0,004
10	+	+	-	+	+	-	-	-	1,007	-0,023	42	-	+	-	+	+	-	+	+	1,246	0,001
11	+	+	-	+	-	+	-	-	1,059	-0,047	43	-	+	-	+	-	+	+	+	1,335	0,008
12	+	+	-	+	-	-	-	+	1,029	-0,018	44	-	+	-	+	-	-	+	-	1,267	0,005
13	+	+	-	-	+	+	+	+	0,939	0,098	45	-	+	-	-	+	+	-	-	1,309	0,009
14	+	+	-	-	+	-	+	+	0,879	0,066	46	-	+	-	-	+	-	-	+	1,246	0,004
15	+	+	-	-	-	+	+	-	0,926	0,095	47	-	+	-	-	-	+	-	+	1,328	0,012
16	+	+	-	-	-	-	+	+	0,871	0,064	48	-	+	-	-	-	-	-	-	1,257	0,006
17	+	-	+	+	+	+	-	-	1,131	0,006	49	-	-	+	+	+	+	+	+	1,424	-0,011
18	+	-	+	+	+	-	-	+	1,066	0,038	50	-	-	+	+	+	-	+	-	1,419	-0,003
19	+	-	+	+	+	-	+	+	1,130	0,007	51	-	-	+	+	-	+	+	-	1,421	-0,013
20	+	-	+	+	-	-	-	-	1,062	0,038	52	-	-	+	+	-	-	+	+	1,429	-0,005
21	+	-	+	-	+	+	+	-	1,047	0,020	53	-	-	+	-	+	+	-	+	1,325	0,001
22	+	-	+	-	+	-	+	+	1,005	0,025	54	-	-	+	-	+	-	-	-	1,313	0,001
23	+	-	+	-	-	+	+	+	1,065	0,026	55	-	-	+	-	-	+	-	-	1,348	-0,001
24	+	-	+	-	-	-	+	-	0,995	0,031	56	-	-	+	-	-	-	-	+	1,328	-0,001
25	+	-	-	+	+	+	+	-	1,313	-0,092	57	-	-	-	+	+	+	-	+	1,460	-0,008
26	+	-	-	+	+	-	+	+	1,252	-0,054	58	-	-	-	+	+	-	-	-	1,441	-0,006
27	+	-	-	+	-	+	+	+	1,305	-0,085	59	-	-	-	+	-	+	-	-	1,517	-0,007
28	+	-	-	+	-	-	+	-	1,245	-0,045	60	-	-	-	+	-	-	-	+	1,536	-0,006
29	+	-	-	-	+	+	-	-	1,260	-0,021	61	-	-	-	-	+	+	+	+	1,438	-0,007
30	+	-	-	-	+	-	-	+	1,164	-0,010	62	-	-	-	-	+	-	+	-	1,429	-0,006
31	+	-	-	-	-	+	-	+	1,268	-0,011	63	-	-	-	-	-	+	+	-	1,506	-0,006
32	+	-	-	-	-	-	-	-	1,158	-0,003	64	-	-	-	-	-	-	+	+	1,466	-0,006

Numa primeira análise dos resultados da tabela 3.1, verifica-se que, na maioria dos ensaios, a eficiência relativa foi superior à unidade. A média geral foi igual a 1,20, indicando que, na média das condições experimentais adotadas, o processo de agregação aumenta a eficiência das estimativas em 20%. As figuras 3.2 e 3.3 apresentam o gráfico de Pareto e o gráfico normal dos efeitos, respectivamente.

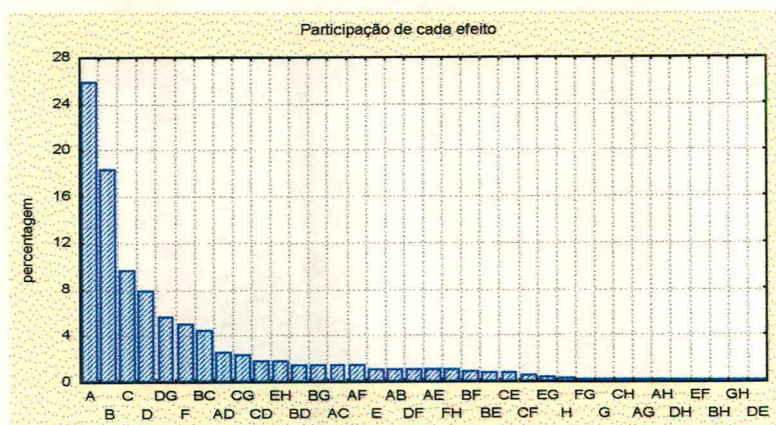


Fig. 3.2 - Participação percentual da magnitude de cada efeito em relação à soma do valor absoluto de todos os efeitos (principais e interações de segunda ordem).

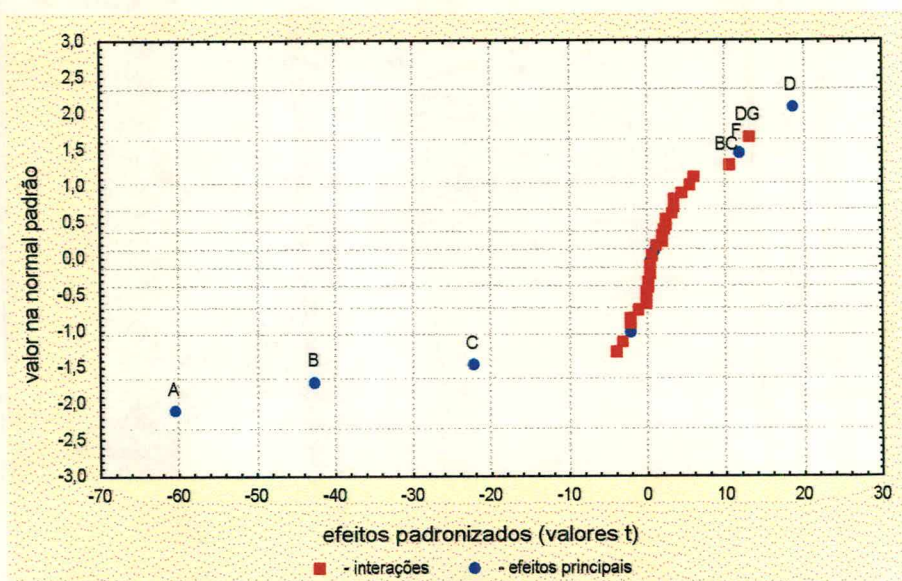


Fig. 3.3 - Gráfico normal dos efeitos (principais e interações de segunda ordem).

A figura 3.2 mostra que os sete primeiros efeitos (A, B, C, D, DG, F e BC) representam, em conjunto, 77% da magnitude total dos efeitos considerados. Além disso, a participação isolada de cada um dos outros efeitos não passa de 3%. A figura 3.3, que apresenta os efeitos numa métrica mais apropriada (segundo a distribuição normal), indica que os efeitos A, B, C e D não devem ser negligenciados. E os efeitos DG, F e BC também podem ser significativos.

Antes de proceder a análise dos efeitos não negligenciados, realizou-se uma análise dos resíduos do modelo ajustado com os sete principais efeitos, conforme mostra a figura 3.4.

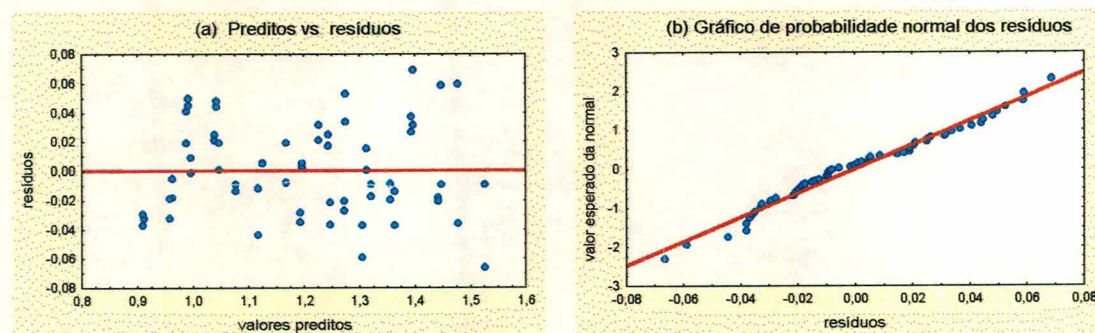


Fig. 3.4 - Análise dos resíduos de um modelo ajustado com os sete maiores efeitos da simulação nº 1.

Os gráficos da figura 3.4 mostram uma boa aderência da curva normal, o que era um resultado já esperado pela forma do cálculo da medida de eficiência. Além disso, não se verificam valores discrepantes, que poderiam resultar de alguma falha na

realização do projeto experimental de simulação. A tabela 3.2 apresenta as estimativas dos sete principais efeitos identificados nas figuras 3.2 e 3.3.

Tabela 3.2 - Estimativa dos principais efeitos na eficiência relativa da agregação.

fator	efeito	erro padrão ¹
média global	1,198	0,002
A	-0,261	0,004
B	-0,184	0,004
C	-0,096	0,004
D	0,079	0,004
F	0,050	0,004
DG (\equiv ADG)	0,056	0,004
BC (\equiv ABC)	0,044	0,004

¹ O erro padrão foi estimado através da agregação dos efeitos de interação de ordem superior a dois.

O fator A é responsável pelo efeito de maior magnitude (ver tabela 3.2), mostrando que, em média, num experimento do tipo 2^6 (nível -1 do fator A), tem-se 26% mais de eficiência com a agregação do que num experimento 2^4 (nível +1 do fator A). O fator B (número de parâmetros no modelo da média) também mostrou-se bastante relevante. A análise destes dois fatores indica que a agregação dos resíduos produz melhores resultados em experimentos com grande número de fatores e poucos parâmetros no modelo da média. São as situações em que se tem grande número de graus de liberdade para a estimação dos resíduos.

O fator C mostra que um número maior de parâmetros no modelo de σ_y^2 também prejudica a eficiência da agregação, embora com menor relevância. O fator F indica que a agregação produz melhores resultados na estimativa dos coeficientes de fatores que produzem fortes efeitos de dispersão. O fator D, colocado na análise para verificar eventuais prejuízos quando o fator suprimido está no modelo de μ_y , mostrou o contrário, isto é, quando o modelo de μ_y inclui o fator que é suprimido no modelo de σ_y^2 , tem-se uma leve melhora com a agregação.

A análise dos efeitos das interações DG e BC ficam prejudicadas por estarem confundidas com efeitos de interação de terceira ordem (ADG e ABC, respectivamente), referentes a fatores com efeitos principais fortes. Assim, não se

pode precisar, com clareza, se os efeitos são causados pelas interações de segunda ou de terceira ordem.

Para se verificar se o processo de agregação pode causar algum vício nas estimativas, fez-se uma análise complementar com os dados do vício (diferença média entre a estimativa com agregação e o valor do parâmetro), constantes na tabela 3.1. Verificou-se que a média global dos vícios não é significativamente diferente de zero e o único fator que provoca alteração significativa, quando considerada uma análise de variância somente com os efeitos principais, é o fator D. O efeito estimado deste fator é pequeno (igual a $-0,03$, que corresponde, em termos relativos, a $4,4\%$). Conclui-se, então, que eventuais vícios nas estimativas, causados pela agregação dos resíduos quadráticos, podem ser negligenciados.

3.2.2 - Experimentos fatoriais 2^k com replicações

Em experimentos com replicações, pode-se também suprimir fatores não relevantes no modelo de σ_y^2 , agregando os resíduos quadráticos médios ou as variâncias amostrais, antes da transformação logarítmica. Nesta seção, procura-se avaliar se tal procedimento aumenta a eficiência das estimativas em experimentos fatoriais 2^k com ênfase no caso de duas replicações.

O projeto experimental para este estudo de simulação usou os fatores que apresentaram efeitos principais significativos na simulação nº 1. Além deles, considerou-se, também, o método de estimação da variância em cada ponto experimental: o método RQ, baseado nos resíduos quadráticos médios, ou o método S2, que utiliza as variâncias amostrais. O quadro 3.2 descreve os cinco fatores que foram usados no estudo de simulação nº 2.

Quadro 3.2 - Fatores e níveis usados no estudo de simulação nº 2.

Fator	Descrição	Nível inferior (-)	Nível superior (+)
A	Tipo de experimento	2^5	2^4
B	Número de parâmetros no modelo da média	Modelo gerado e especificado no processo de estimação somente com o termo constante ($\beta_0 = 0$) e efeitos principais ($\beta_i = 1$, $i = 1, 2, \dots, k-1$ e β_k definido em D).	Incluem-se interações de segunda ordem ($\beta_{ij} = 0,5$, $i < j = 1, 2, 3$ no 2^4 e $\beta_{ij} = 0,5$, $i < j = 1, 2, 3, 4$ no 2^5).

continua...

Quadro 3.2 - Fatores e níveis usados no estudo de simulação nº 2 (continuação).

Fator	Descrição	Nível inferior (-)	Nível superior (+)
C	Número de parâmetros no modelo da variância	Modelo gerado e especificado no processo de estimação somente com θ_0 e θ_1 , sendo $\theta_0 = 0$ e θ_1 especificado em F.	Incluem-se os termos $\theta_2 = \theta_3 = 0,347$ no 2^4 e os termos $\theta_2 = \theta_3 = \theta_{12} = \theta_{23} = 0,347$ no 2^5 .
D	Efeito na média do fator suprimido no modelo da variância	Efeito nulo ($\beta_k = 0$) e excluído no processo de estimação.	$\beta_k = 1$ e incluído na especificação do modelo para a estimação.
F	Força do efeito de dispersão	$\theta_1 = 0,549$, que altera a variância em 3 vezes.	$\theta_1 = 0,896$, que altera a variância em 6 vezes.
M	Método de estimação das variâncias de cada ponto experimental	RQ (resíduo quadrático médio)	S2 (variância amostral)

Realizou-se o estudo de simulação com base num projeto fatorial completo 2^6 , onde os resultados da eficiência relativa da agregação, baseados em 2.000 realizações simuladas em cada condição experimental, são apresentados na tabela 3.3. A tabela 3.3 apresenta somente as combinações de níveis para o método RQ ($M = -1$), pois no método S2 os resultados foram constantes para cada tipo de experimento, conforme aparece na última linha da tabela.

Tabela 3.3 - Resultados da eficiência relativa com a agregação. Estudo de simulação nº 2.

	A	B	C	D	F	ef.		A	B	C	D	F	ef.
M=-1:	-	-	-	-	-	1,255	M=-1:	+	-	-	-	-	1,197
	-	-	-	-	+	1,257		+	-	-	-	+	1,223
	-	-	-	+	-	1,287		+	-	-	+	-	1,235
	-	-	-	+	+	1,275		+	-	-	+	+	1,277
	-	-	+	-	-	1,238		+	-	+	-	-	1,170
	-	-	+	-	+	1,245		+	-	+	-	+	1,192
	-	-	+	+	-	1,253		+	-	+	+	-	1,238
	-	-	+	+	+	1,254		+	-	+	+	+	1,262
	-	+	-	-	-	1,210		+	+	-	-	-	1,119
	-	+	-	-	+	1,216		+	+	-	-	+	1,133
	-	+	-	+	-	1,217		+	+	-	+	-	1,172
	-	+	-	+	+	1,226		+	+	-	+	+	1,195
	-	+	+	-	-	1,192		+	+	+	-	-	1,106
	-	+	+	-	+	1,204		+	+	+	-	+	1,125
	-	+	+	+	-	1,198		+	+	+	+	-	1,178
	-	+	+	+	+	1,209		+	+	+	+	+	1,184
M=+1:	-	*	*	*	*	1,534	M=+1:	+	*	*	*	*	1,453

Observa-se na tabela 3.3 que em todos os ensaios a eficiência relativa com a agregação, na estimação de θ_1 , foi superior à unidade e, em especial no método S2, o ganho em eficiência foi de 53% no projeto 2^5 e 45% no projeto 2^4 ,

independentemente dos outros fatores. De certa forma este resultado já era esperado, pois, no método S2, a variância do estimador independe da parte determinística dos modelos (ver Nair e Pregibon, 1988). No método RQ, o ganho em eficiência foi mais modesto: 21%, em média. A figura 3.5 mostra o gráfico normal dos efeitos com o fator M fixado em -1 (método RQ).

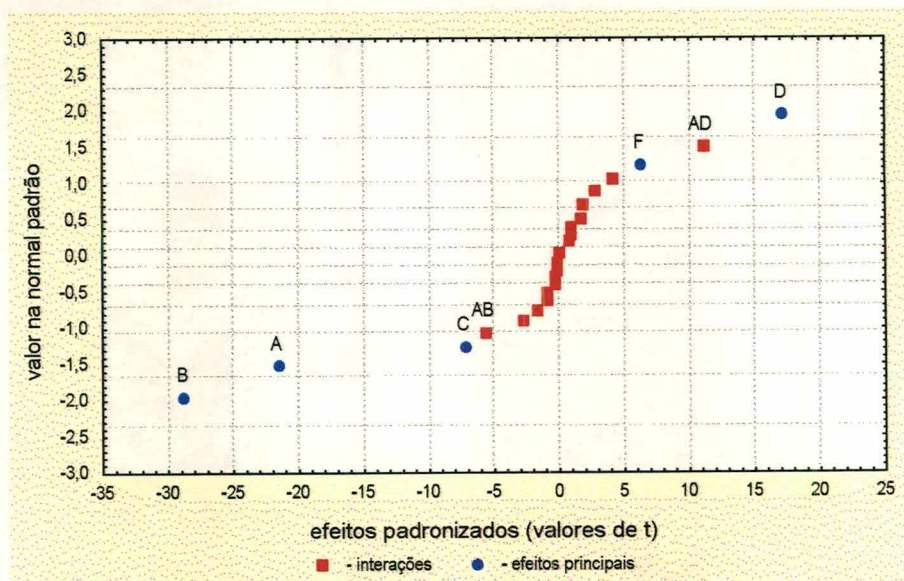


Fig. 3.5 - Gráfico normal dos efeitos dos dados da simulação nº 2, método RQ.

A figura 3.5 mostra que os fatores A, B e D, com efeitos estimados de $-0,05$, $-0,06$ e $0,04$, respectivamente, são os que mais alteram o desempenho do procedimento da agregação. Ou seja, este procedimento, sob o método RQ, produz melhores resultados em experimentos com grande número de fatores e poucos parâmetros no modelo da média, confirmando a tendência verificada na simulação nº 1. Confirma-se, também, que se o fator suprimido no modelo da variância é um fator ativo no modelo da média ($D = +1$), o resultado da agregação torna-se melhor.

Para se avaliar o efeito da agregação em experimentos com mais observações em cada ponto experimental, repetiu-se a simulação nº 2 num projeto experimental 2^5 com 4 replicações. O método S2 produziu novamente um ganho em eficiência constante e igual a 17%. E o método RQ apresentou um ganho médio em eficiência de 12%. Conclui-se, então, que a agregação ainda produz resultados melhores em experimentos com quatro replicações, porém com ganho em eficiência bem inferior aos experimentos com duas replicações.

3.2.3 - Experimentos fatoriais 3^k

Uma complementação do estudo foi feita com experimentos fatoriais do tipo 3^k , onde cada um dos k fatores são observados em três níveis, normalmente codificados como -1, 0 e 1. Utilizou-se $k = 3$, uma e duas replicações e respostas simuladas com distribuição normal de média e variância dadas pelas seguintes expressões:

$$\mu_y = x_1 + x_2 + x_3 + 0,5(x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3) \quad (3.10)$$

$$\sigma_y^2 = \exp\{0,896x_1 + 0,549x_1^2\} \quad (3.11)$$

Para o processo de estimação, os modelos foram especificados por

$$\mu_y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{33}x_3^2 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 \quad (3.12)$$

$$\sigma_y^2 = \exp\{\theta_0 + \theta_1x_1 + \theta_{11}x_1^2\} \quad (3.13)$$

No experimento simulado sem replicação, a estimação foi realizada pelo método RQ e, no experimento simulado com duas replicações, a estimação foi realizada pelos métodos S2 e RQ. Assim como nos casos anteriores, o método RQ foi usado com duas iterações. A tabela 3.4 apresenta a eficiência da agregação, na estimação dos parâmetros θ_1 e θ_{11} , baseando-se em 2.000 realizações simuladas de cada experimento.

Tabela 3.4 - Resultados da eficiência relativa com a agregação em experimentos do tipo 3^3 .

nº de replicações	método de estimação	eficiência relativa na estimação	
		θ_1	θ_{11}
1	RQ	1,32	1,25
2	RQ	1,36	1,31
	S2	1,82	1,71

Os resultados apresentados na tabela 3.4 mostram que o processo de agregação também tende a produzir melhores estimativas na classe de experimentos 3^k , com ganhos maiores de eficiência sob o método S2.

3.3 - A TRANSFORMAÇÃO LOGARÍTMICA MODIFICADA

Para se usarem os resíduos quadráticos médios ($\hat{\epsilon}_i^2$) ou as variâncias amostrais (S_i^2) como se fossem observações em um modelo de regressão, torna-se necessária uma transformação para estabilizar as suas variâncias e tornar as estatísticas com distribuição aproximadamente normal. A transformação logarítmica, $\log(S_i^2)$ ou $\log(\hat{\epsilon}_i^2)$, garante totalmente a primeira condição e atinge parcialmente a segunda quando o número de replicações é grande. Por outro lado, quando o número de replicações é pequeno, o que é comum na prática, pode-se ter S_i^2 (ou $\hat{\epsilon}_i^2$) próximos de zero meramente por acaso. Neste caso, a transformação provoca valores discrepantes, os quais prejudicam o processo de estimação. Estes valores discrepantes, por serem devido a pontos muito próximos da média da resposta, são chamados de *inliers*.

Um modelo linear especificado em termos do logaritmo da variância da resposta tem sido a prática mais comum (Harvey, 1976, Box e Meyer, 1986, Nair e Pregibon, 1988, Engel, 1992, Greco, 1993, Verbyla, 1993, Myers e Montgomery, 1995 e Hughes-Oliver et al., 1998). Conforme ponderam Nelder e Lee (1998), a função de ligação da variância com um modelo linear não precisa ser necessariamente a logarítmica, mas normalmente se tem poucos dados para se avaliar empiricamente outras funções. Outra vantagem da utilização do logaritmo é que muitas vezes ele induz aditividade, permitindo uma modelagem mais simples, conforme argumentam Bartlett e Kendall (1946) e Hamada e Nelder (1997).

Na estimação por MQO, vários autores sugerem algum procedimento para reduzir o problema causado pelos *inliers*, como a eliminação dos resíduos (ou variâncias amostrais) próximos de zero ou a utilização de métodos de estimação robusta (Carroll e Ruppert, 1988, Ferrer e Homero, 1995 e Engel e Huele, 1996). Observa-se que, na análise de dados experimentais, a eliminação de observações compromete as propriedades do projeto utilizado.

O presente trabalho sugere uma alteração na transformação, adicionando-se uma quantidade positiva q em todas as variâncias amostrais (ou resíduos quadráticos

médios), antes da aplicação do logaritmo. Assim, para o método S2, a equação da variância passa a ter a forma

$$\log(S_i^2 + q) = z_i' \theta + u_i, \quad i = 1, 2, \dots, M \quad (3.14)$$

onde: S_i^2 são as variâncias amostrais nos pontos experimentais ensaiados;

z_i são as linhas da matriz de planejamento Z e

u_i são os erros aleatórios.

Este procedimento será referenciado como transformação logarítmica modificada. Após a estimação do vetor de parâmetros θ por MQO, pode-se expressar a equação de predição das variâncias por

$$\hat{\sigma}_i^2 = \exp\{z_i' \hat{\theta}\} - q, \quad i = 1, 2, \dots, M \quad (3.15)$$

onde $\hat{\theta}$ é a estimativa do vetor θ . A figura 3.6 ilustra a função logarítmica agindo sobre a distribuição de S^2 e de $S^2 + q$, onde observa-se que a adição da constante q evita a região onde a função logarítmica assume valores extremamente baixos.

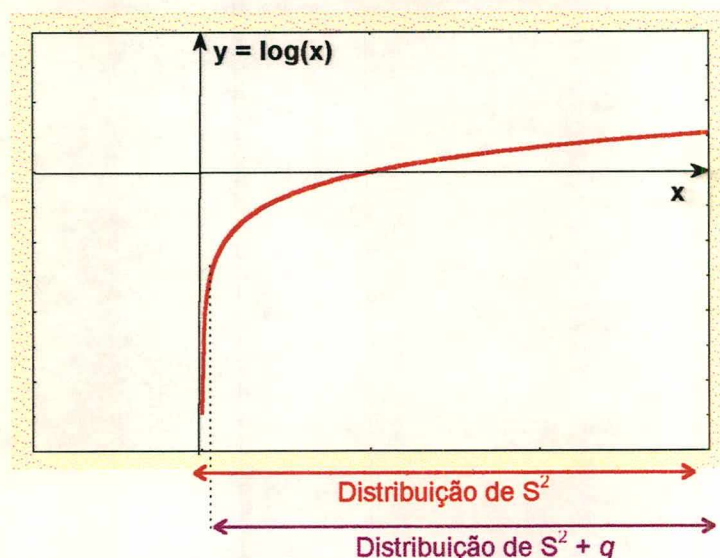


Fig. 3.6 - A função logarítmica e a base das distribuições de S^2 e de $S^2 + q$.

A transformação proposta funciona como uma espécie de estimação robusta, pois reduz a distância dos *inliers* em relação ao centro da distribuição. Vining e Bohn (1998) usaram esta transformação com $q = 1$, justificando-a em termos da aplicação que continha variâncias amostrais nulas, impossibilitando a transformação logarítmica padrão.

Para se levar em conta a magnitude das variâncias, propõe-se, no presente trabalho, fazer a quantidade q como função da magnitude da média das variâncias amostrais, da seguinte forma:

$$q = c \bar{S}^2 \quad (3.16)$$

onde: c é uma constante positiva e

$$\bar{S}^2 = \frac{1}{M} \sum_{i=1}^M S_i^2 \quad \text{é a média de todas as variâncias amostrais.}$$

No método RQ, a transformação pode ser feita de forma análoga, substituindo-se as variâncias amostrais pelos resíduos quadráticos médios, ou seja, estabelece-se a equação de regressão por $\log(\hat{e}_i^2 + q) = z_i' \theta + u_i$, onde $q = c \bar{e}^2$.

Um cuidado que se deve ter na transformação proposta é que o modelo pode prever valores negativos para algumas variâncias. Neste trabalho, condicionou-se como valor mínimo das variâncias preditas a quantidade correspondente a 0,5% da média global das variâncias amostrais (ou da média global dos resíduos quadráticos, no caso do método RQ). No capítulo 5 será apresentada outra forma de se estabelecer a equação de predição das variâncias.

3.3.1 - A distribuição dos erros na equação de regressão proposta

O problema da criação de valores discrepantes, devido à transformação, costuma ser bastante grave quando se trabalha com os resíduos quadráticos individuais ou com variâncias amostrais calculadas com apenas duas observações. Nestes casos, a distribuição dos erros aproxima-se proporcionalmente de uma distribuição qui-quadrado com um grau de liberdade (χ_1), que é muito assimétrica e com alta densidade de probabilidade sobre valores positivos próximos de zero. É sobre esta distribuição que se fará a primeira análise da equação de regressão proposta.

Distribuições proporcionais à χ_1 podem ser obtidas a partir do quadrado de distribuições normais de média zero e variâncias σ_i^2 , $N(0, \sigma_i^2)$, $i = 1, 2, \dots$. Estas distribuições serão representadas por Z_i^2 , $i = 1, 2, \dots$ e terão variâncias $\text{Var}(Z_i^2) = 2\sigma_i^4$, $i = 1, 2, \dots$ (Carroll e Ruppert, 1988, p. 78). Conforme demonstrado por Bartlett e

Kendall (1946) e por Harvey (1976), $\text{Var}\{\log(Z_1^2)\} = \text{Var}\{\log(Z_2^2)\} = \dots = 4,93$, ou seja, a transformação logarítmica estabiliza a variância. O objetivo, nesta seção, é verificar como ficam as variâncias e outras características da distribuição, quando a transformação for do tipo

$$\log(Z_i^2 + c\bar{Z}^2), \text{ onde } 0 \leq c < 1 \text{ e } \bar{Z}^2 = \frac{\sum_{i=1}^n Z_i^2}{n} \quad (3.17)$$

O estudo foi realizado com amostras simuladas de distribuições normais de variâncias 1 e 9, que produzem, quando elevadas ao quadrado, amostras de distribuições χ_1 com variâncias 2 e 162, respectivamente. Foram usados na transformação os seguintes valores de c : 0, 0,01 e 0,10. Segue a descrição da simulação, cujo programa computacional encontra-se no apêndice.

Realizaram-se 10.000 observações da $N(0, 1)$ e 10.000 observações da $N(0, 9)$, fazendo-se uso da subrotina RANNOR do SAS. Tomou-se o quadrado destas observações, obtendo-se dois conjuntos de 10.000 valores, os quais podem ser considerados como amostras de distribuições proporcionais à distribuição χ_1 . Estes valores foram denominados de Z_i^2 , $i = 1, 2, \dots, 20.000$. Calculou-se a média aritmética dos 20.000 valores (\bar{Z}^2) e consideraram-se proporções c desta média ($c = 0, 0,01$ e $0,10$). A tabela 3.5 apresenta as variâncias de $\log(Z_i^2 + c\bar{Z}^2)$, calculadas em cada um dos conjuntos de 10.000 observações.

Tabela 3.5 - Variâncias de $\log(Z_i^2 + c\bar{Z}^2)$, segundo a população geradora e valores de c .

população geradora	valor de c		
	0 ¹	0,01	0,10
$N(0, 1)$	4,89 (4,93)	1,80	0,49
$N(0, 9)$	4,87 (4,93)	3,18	1,74
Razão	0,99 (1,00)	1,77	3,55

¹ Os valores entre parênteses referem-se ao resultado teórico de Harvey (1976).

Observa-se, nos resultados da tabela 3.5, que a transformação proposta deixa de garantir a propriedade de estabilização da variância. Contudo, considerando que uma diferença de 9 vezes nas variâncias das observações acarreta uma diferença de 81 vezes na variância dos desvios quadráticos, pode-se dizer que a transformação

proposta tem razoável sucesso na estabilização da variância, já que, na situação mais extrema ($c = 0,10$), as variâncias são alteradas por um fator em torno de 3,55. Além do mais, as variâncias produzidas pela transformação proposta são bastante inferiores à variância estabilizada pela transformação logarítmica padrão, podendo acarretar num ganho em eficiência no processo de estimação dos parâmetros do modelo de regressão.

Resultados análogos foram obtidos usando como populações geradoras $N(0, 10)$ e $N(0, 90)$, indicando que a magnitude das variâncias não é relevante, mas somente a proporcionalidade entre elas. A tabela 3.6 mostra outras características da distribuição de $\log(Z_i^2 + c \bar{Z}^2)$, considerando a distribuição $N(0, 1)$ como a população geradora, num processo de simulação com 10.000 observações.

Tabela 3.6 - Algumas características da distribuição de $\log(Z_i^2 + c \bar{Z}^2)$

Características da distribuição	valor de c		
	0 ¹	0,01	0,10
coef. de assimetria	-1,51 (-1,54)	-0,43	0,19
coef. de curtose	3,62 (4,00)	-0,69	-0,99

¹ Os valores entre parêntesis referem-se aos resultados teóricos de Harvey (1976).

Os resultados da tabela 3.6 mostram que as distribuições de $\log(Z_i^2 + c \bar{Z}^2)$, com $c = 0,01$ e $c = 0,10$, são mais próximas da distribuição normal do que a distribuição de $\log(Z_i^2)$, cujos resultados estão na coluna referente a 0%. A figura 3.7 ilustra a distribuição de $\log(Z_i^2 + c \bar{Z}^2)$, com $c = 0, 0,01$ e $0,10$.

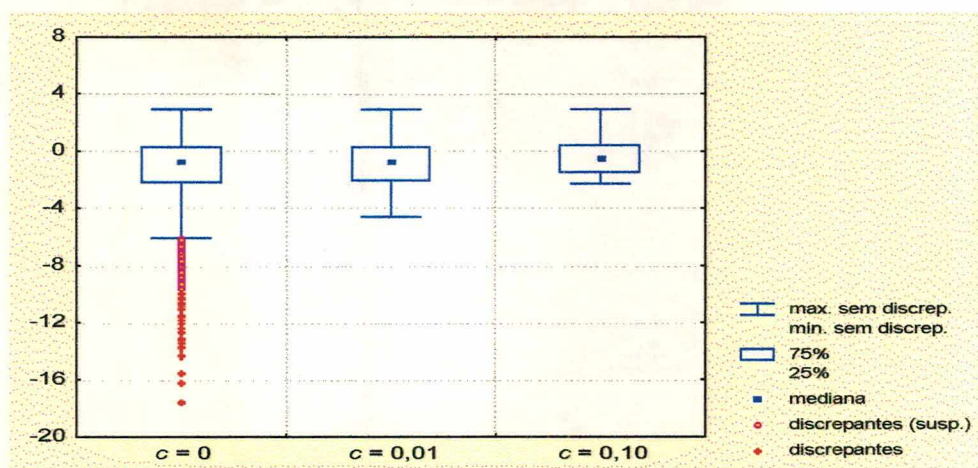


Fig. 3.7 - Diagramas de caixas de 10.000 observações simuladas da distribuição de $\log(Z_i^2 + c \bar{Z}^2)$, com $c = 0, 0,01$ e $0,10$.

Observa-se na figura 3.7 que a transformação logarítmica modificada gera uma distribuição com dispersão menor do a transformação logarítmica padrão e, também, elimina o problema dos pontos discrepantes.

3.3.2 - Fatores relevantes na eficiência da transformação proposta

Para verificar, em termos práticos, a adequação da transformação proposta, realizou-se um novo estudo de simulação sobre diferentes condições experimentais. Incluiu-se, neste estudo, nove fatores, os quais são descritos no quadro 3.3.

Quadro 3.3 - Fatores e níveis usados no estudo de simulação nº 3.

Fator	Descrição	Nível inferior (-)	Nível superior (+)
A	Tipo de experimento	Fatorial 2^5 , gerando 32 pontos experimentais.	CCD com 4 fatores e 31 pontos experimentais (16 pontos fatoriais, 8 axiais e 7 centrais).
B	Número de parâmetros no modelo da média	Modelo formado somente com o termo constante e efeitos principais lineares (6 parâmetros no 2^5 e 5 parâmetros no CCD).	Incluem-se mais 10 parâmetros (interações de segunda ordem e, no caso do CCD, termos quadráticos); os valores dos parâmetros são descritos pelo fator E.
C	Número de parâmetros no modelo da variância	2 parâmetros: termo constante e θ_1 , onde o valor de θ_1 é definido em F.	Incluem-se mais 14 parâmetros, sendo todos gerados com o valor 0,347, que altera a variância por um fator de dois.
D	Número de replicações	2 (totalizando 64 e 62 observações no 2^5 e CCD, respectivamente)	4 (totalizando 128 e 124 observações no 2^5 e CCD, respectivamente)
E	Efeitos de locação	$\beta_i = 1$ ($i = 0, 1, 2, \dots$) $\beta_{ij} = 0,5$ ($i \leq j = 0, 1, \dots$)	$\beta_i = 8$ ($i = 0, 1, 2, \dots$) $\beta_{ij} = 2$ ($i \leq j = 0, 1, \dots$)
F	Força do efeito de dispersão	$\theta_1 = 0,549$, que altera a variância em 3 vezes.	$\theta_1 = 0,896$, que altera a variância em 6 vezes.
G	Magnitude geral das variâncias	$\sigma_0^2 = 1$	$\sigma_0^2 = 3$
M	Método de estimação das variâncias em cada ponto experimental	RQ	S2
W	Porcentagem da variância global adicionada a cada estimativa da variância (valor de c)	Cinco níveis: -2, -1, 0, 1, 2 (correspondendo aos valores percentuais: $\frac{1}{2}$, 1, 2, 4 e 8, respectivamente).	

No fator A, tipo de experimento, procurou-se fazer com que a diferença entre os níveis fique por conta da disposição dos pontos experimentais e não em relação ao número de observações como no estudo de simulação nº 1. Inclui-se, no fator D, o número de replicações. Os fatores B, C, E, F, G e M são análogos aos dos estudos de simulação anteriores.

O fator W, que especifica a forma da transformação proposta, foi tomado em cinco níveis para permitir avaliar a qualidade de um ajuste de uma função polinomial de segundo grau. A relação dos níveis com os valores de c foi feita em escala logarítmica porque se verificou, em ensaios preliminares, que o ajuste de uma função quadrática torna-se adequado sob esta escala.

No experimento central composto de segunda ordem ($A = +1$), os cálculos das variâncias amostrais e dos resíduos quadráticos médios foram feitos sempre com o mesmo número de observações. Ou seja, no ponto central, onde existem sete vezes mais observações, calcularam-se sete estatísticas. Este procedimento também foi usado em Castilho (1996). Assim como nas simulações anteriores, em todos os ensaios com $H = -1$ (método RQ), realizou-se o processo de estimação com 2 iterações e, também, considerou-se a quantidade 0,5% da média de todos os resíduos quadráticos como o valor mínimo de predição da variância. As predições das variâncias são usadas para se estimar os parâmetros do modelo da média por MQG e, com isto, obter novos resíduos (processo iterativo).

Para a realização do estudo, utilizou-se um projeto experimental $5 \times 2^{8-2}$. O 5 refere-se aos níveis do fator W, os quais são cruzados com todos os níveis dos outros fatores, que estão dispostos num experimento fatorial fracionado 2^{8-2} . Desta forma, pode-se avaliar as interações do fator W com todos os outros fatores em estudo. As combinações da parte 2^{8-2} foram feitas da mesma forma do estudo de simulação nº 1 (ver tabela 3.1), sendo que em cada combinação de níveis apresentada na tabela 3.1, realizaram-se cinco ensaios, conforme o nível de W: -2, -1, 0, 1 e 2. A tabela 3.7 apresenta, para cada ensaio, os resultados da eficiência da transformação logarítmica modificada em relação à transformação logarítmica padrão, na estimação de θ_1 . Cada resultado baseia-se em 1.000 realizações simuladas do experimento.

Usou-se a mesma sequência de números aleatórios em todos os ensaios. O programa computacional encontra-se no apêndice.

Tabela 3.7 - Resultados da eficiência relativa da transformação logarítmica modificada no estudo de simulação nº 3.

ref. do ensaio ¹	eficiência para cada nível de W					ref. do ensaio ¹	eficiência para cada nível de W				
	-2	-1	0	1	2		-2	-1	0	1	2
1	1,19	1,25	1,18	0,88	0,53	33	1,02	0,70	0,41	0,23	0,14
2	1,12	1,20	1,31	1,39	1,29	34	1,33	1,39	1,28	0,95	0,63
3	1,17	1,21	1,13	0,86	0,53	35	1,02	0,73	0,44	0,26	0,16
4	1,12	1,21	1,33	1,42	1,33	36	1,31	1,38	1,25	0,91	0,58
5	1,54	1,66	1,66	1,49	1,17	37	1,41	1,21	0,91	0,64	0,45
6	1,66	1,87	2,11	2,34	2,44	38	2,05	2,29	2,42	2,34	2,05
7	1,65	1,72	1,70	1,53	1,25	39	1,63	1,47	1,20	0,93	0,70
8	1,47	1,67	1,92	2,18	2,31	40	1,76	1,96	2,05	1,89	1,54
9	1,10	1,17	1,24	1,20	0,88	41	1,08	1,13	1,17	1,06	0,72
10	1,06	1,11	1,18	1,26	1,27	42	1,06	1,11	1,18	1,24	1,20
11	1,10	1,15	1,19	1,11	0,82	43	1,09	1,14	1,15	1,02	0,70
12	1,05	1,09	1,17	1,26	1,29	44	1,05	1,08	1,14	1,21	1,18
13	1,25	1,37	1,50	1,55	1,38	45	1,18	1,28	1,37	1,34	1,09
14	1,53	1,70	1,91	2,14	2,30	46	1,53	1,70	1,90	2,08	2,15
15	1,54	1,65	1,71	1,65	1,43	47	1,52	1,61	1,63	1,50	1,23
16	1,16	1,25	1,39	1,58	1,75	48	1,14	1,22	1,35	1,50	1,61
17	1,17	1,21	1,13	0,86	0,53	49	1,02	0,73	0,44	0,26	0,16
18	1,10	1,17	1,27	1,32	1,18	50	1,29	1,35	1,20	0,85	0,53
19	1,15	1,19	1,10	0,79	0,46	51	1,00	0,68	0,38	0,21	0,13
20	1,12	1,20	1,31	1,39	1,29	52	1,33	1,39	1,28	0,95	0,63
21	1,65	1,72	1,70	1,53	1,25	53	1,63	1,47	1,20	0,93	0,70
22	1,32	1,46	1,64	1,81	1,83	54	1,57	1,73	1,75	1,54	1,18
23	1,40	1,49	1,46	1,24	0,90	55	1,27	1,05	0,75	0,51	0,34
24	1,66	1,87	2,11	2,34	2,44	56	2,05	2,29	2,42	2,34	2,05
25	1,10	1,15	1,19	1,11	0,82	57	1,09	1,14	1,15	1,02	0,70
26	1,04	1,08	1,14	1,20	1,19	58	1,04	1,08	1,14	1,19	1,14
27	1,08	1,13	1,16	1,08	0,77	59	1,07	1,11	1,13	1,00	0,66
28	1,06	1,11	1,18	1,26	1,27	60	1,06	1,11	1,18	1,24	1,20
29	1,54	1,65	1,71	1,65	1,43	61	1,52	1,61	1,63	1,50	1,23
30	1,12	1,20	1,31	1,44	1,54	62	1,12	1,19	1,28	1,38	1,39
21	1,17	1,25	1,33	1,31	1,10	63	1,14	1,20	1,23	1,13	0,86
32	1,53	1,70	1,91	2,14	2,30	64	1,53	1,70	1,90	2,08	2,15

¹ Refere-se aos ensaios da tabela 3.1, mas, aqui, correspondem a cinco ensaios, segundo os níveis do fator W.

A figura 3.8 apresenta o gráfico normal dos efeitos. Os pontos do gráfico referem-se aos efeitos principais e de interações de segunda ordem. No caso do fator W, estudou-se o seu efeito linear (W), quadrático (W^2) e as suas interações com os demais fatores. Dentre os efeitos não negligenciáveis, o gráfico identifica os efeitos principais e as interações com o fator W.

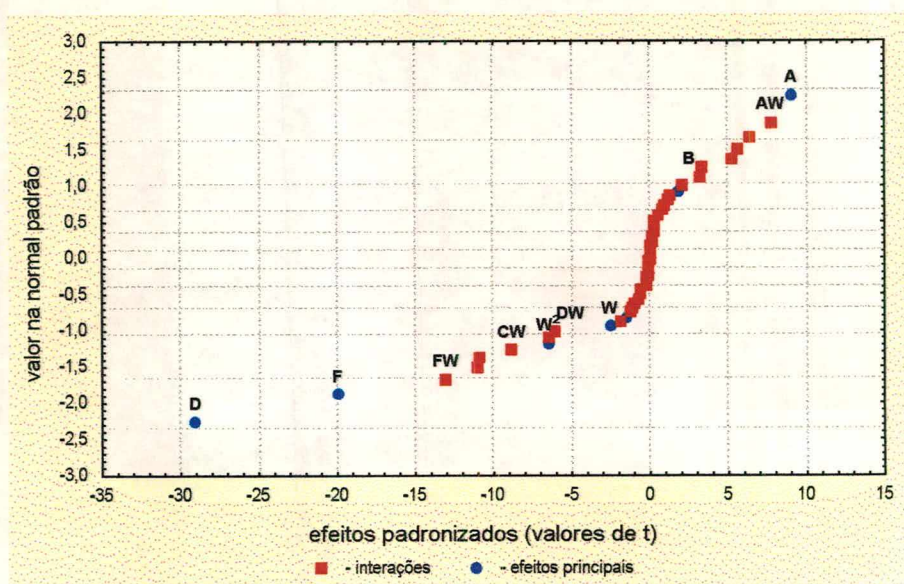


Fig. 3.8 - Gráfico normal dos efeitos do estudo de simulação nº 3.

A tabela 3.8 apresenta os maiores efeitos do estudo de simulação nº 3. Os erros padrão das estimativas foram calculados a partir de uma pseudo estimativa da variância dos erros, baseada na agregação dos efeitos de interações de ordem superior a dois.

Tabela 3.8 - Estimativa dos efeitos significativos do estudo de simulação nº 3.

fator	efeito	erro padrão	fator	efeito	erro padrão
média global	1,290	0,009	AW	0,099	0,012
A	0,164	0,018	BE	0,059	0,018
B	0,057	0,018	BG	0,037	0,018
D	-0,526	0,018	CD	-0,116	0,018
F	-0,362	0,018	CF	-0,199	0,018
W	-0,066	0,021	CW	-0,113	0,012
W ²	-0,140	0,018	DF	0,094	0,018
AC	0,101	0,018	DW	-0,078	0,012
AF	0,115	0,018	EF	0,057	0,018
AH	-0,201	0,018	FW	-0,167	0,012

A tabela 3.8 mostra que, em média, a transformação proposta é superior à transformação logarítmica padrão (ganho em eficiência de 29%). Contudo, vários fatores agem fortemente sobre esta resposta. Em primeiro lugar, o fator D (número de replicações) age de forma inversa, mostrando que a transformação proposta gera um ganho em eficiência muito maior quando se tem experimentos com apenas duas replicações. O resultado acerca do fator F indica que a transformação proposta é mais interessante para os coeficientes de fatores que provocam alterações pequenas na

variância do processo. Já o valor positivo no fator A indica que a transformação proposta produz melhores resultados em experimentos do tipo CCD. O desempenho relativo do procedimento proposto também aumenta quando se tem poucos graus de liberdade para os resíduos ($B = +1$).

Algumas interações também mostraram efeitos bastante fortes, como o caso de AH e CF. O primeiro caso indica que quando se está trabalhando com planejamento do tipo 2^5 , os resultados da transformação proposta são melhores sob o método S2, enquanto que em planejamento do tipo CCD, melhores resultados são obtidos sob o método RQ. A interação CF mostra que quando se está estimando um efeito de dispersão forte, a transformação proposta produz melhor eficiência se não houver muitos outros parâmetros de dispersão a serem estimados. Para a estimação de efeitos de dispersão fracos, a relação é inversa.

Os principais resultados deste estudo de simulação, contudo, dizem respeito ao fator W, que é um fator controlável no processo de estimação dos parâmetros do modelo. O valor negativo e significativo do coeficiente de W^2 indica que existe um valor ótimo de W na região de investigação, porém as interações do fator W com os outros fatores sugerem que este nível ótimo de W depende do nível dos outros fatores estudados. Esta análise é apresentada nas figuras 3.9 e 3.10, que ilustram o ajuste da equação de regressão, enfatizando os fatores W, A, C, D e F. A região acima da linha verde indica maior eficiência para a transformação proposta.

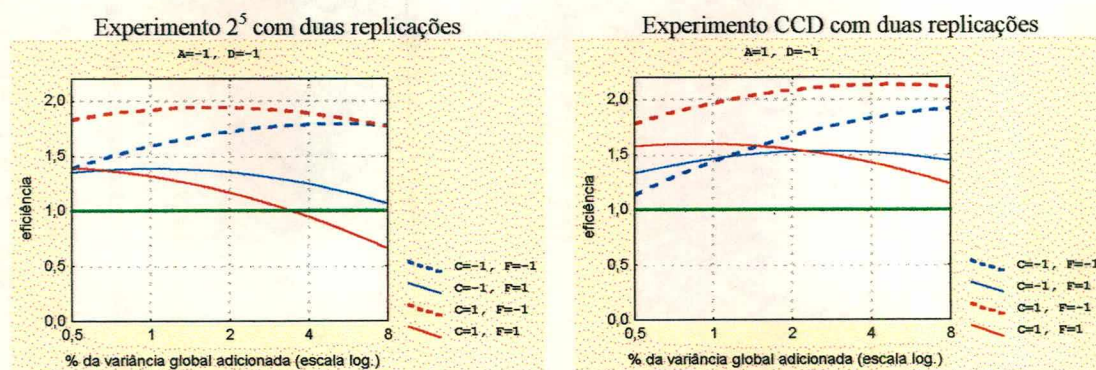


Fig. 3.9 - Eficiência da transformação proposta em função da percentagem da variância global adicionada às estatísticas. Experimentos com duas replicações.

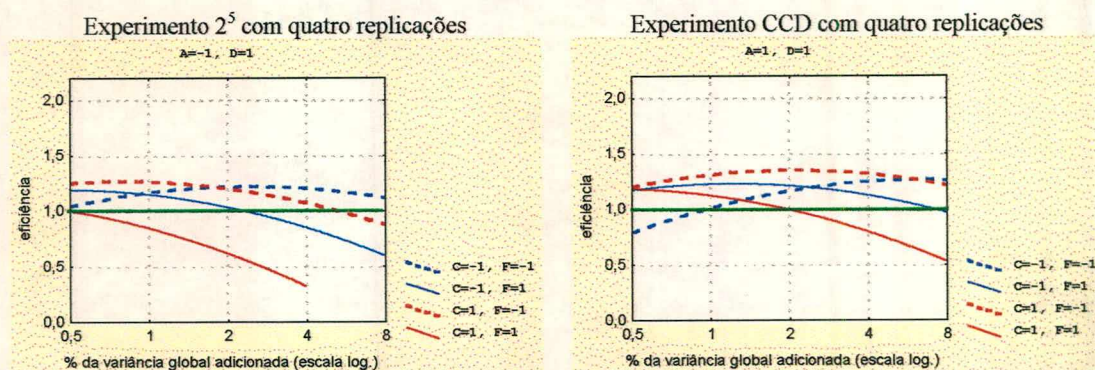


Fig. 3.10 - Eficiência da transformação proposta em função da percentagem da variância global adicionada às estatísticas. Experimentos com quatro replicações.

Observando as figuras 3.9 e 3.10, verifica-se que o ganho de eficiência com a transformação proposta, assim como o valor de c da transformação, dependem bastante dos fatores em estudo. Nos experimentos com duas replicações (figura 3.9), o ganho de eficiência com o procedimento proposto é bastante grande e o valor a ser escolhido para c depende, em grande parte, da força dos efeitos de dispersão. Um valor de c em torno de 4% parece ser adequado quando os efeitos de dispersão forem pequenos ou moderados ($F = -1$). Por outro lado, se houver efeitos de dispersão elevados ($F = +1$), deve-se escolher c em torno de 1%. Já nos experimentos com quatro replicações, só se deve usar a transformação proposta se os efeitos de dispersão não forem muito elevados. E, por segurança, o valor de c não deve ser muito grande (no máximo, $c = 2\%$).

No estudo de simulação nº 3, observou-se o efeito da transformação no processo de estimação do parâmetro de maior magnitude do modelo da variância. Realizam-se, a seguir, alguns ensaios adicionais com a análise ampliada a todos os parâmetros. Estes ensaios adicionais foram feitos com experimentos não replicados nos pontos fatoriais e a estimação pelo método RQ com duas iterações.

Consideraram-se experimentos do tipo 2^4 , 2^5 e CCD com 4 fatores e 7 replicações no ponto central. As equações das médias dos experimentos fatoriais foram especificadas como de primeiro grau com coeficientes unitários, enquanto a equação da média do CCD foi especificada como de segundo grau, com coeficientes unitários nos termos lineares e iguais a 0,5 nos termos quadráticos e de interações. As equações das variâncias dos experimentos fatoriais foram especificadas por

$$\sigma_y^2 = \exp\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3\} \quad (3.18)$$

com $\theta_0 = 0$, $\theta_1 = 0,896$, $\theta_2 = 0,693$ e $\theta_3 = 0,347$. No CCD, a equação da variância foi especificada por

$$\sigma_y^2 = \exp\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1^2\} \quad (3.19)$$

com $\theta_0 = 0$, $\theta_1 = 0,896$, $\theta_2 = 0,693$ e $\theta_{11} = 0,347$.

Em cada experimento, simulou-se 2.000 amostras. Para cada amostra simulada, aplicou-se a técnica dos mínimos quadrados sobre os resultados da transformação $\log(\hat{e}_i^2)$ e, também, sobre os resultados de $\log(\hat{e}_i^2 + c\bar{e}^2)$, com c variando de 0,005 a 0,64, segundo uma progressão geométrica de razão 2. Para cada valor de c , comparou-se os dois tipos de transformação logarítmica (padrão e modificada), em termos da eficiência na estimação de θ_1 , θ_2 e θ_3 (θ_{11}). A figura 3.11 apresenta os resultados.

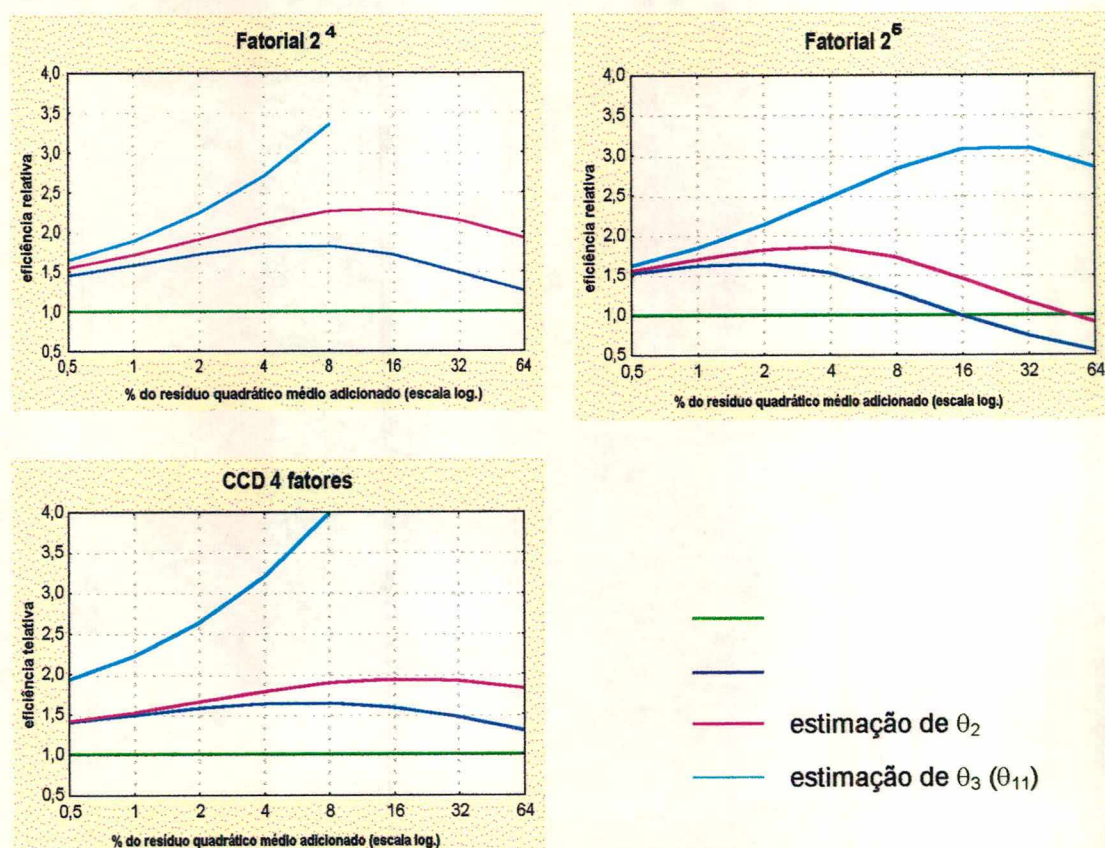


Fig. 3.11 - Eficiência da transformação proposta em função de c e do parâmetro do modelo da variância em três tipos de experimentos.

Observa-se, na figura 3.11, que nos três projetos de experimentos analisados, a transformação logarítmica modificada pode produzir grandes ganhos de eficiência

quando comparada à transformação logarítmica padrão. Contudo, o valor ideal de c , assim como o ganho em eficiência, variam de acordo com a magnitude do parâmetro que está sendo estimado. Nos parâmetros de menor magnitude, tem-se condições mais favoráveis à transformação proposta.

3.3.3 - A transformação em experimentos fatoriais do tipo 2^k

Nesta seção, busca-se uma forma de se estabelecer adequadamente a transformação logarítmica modificada em experimentos fatoriais do tipo 2^k , completos ou fracionados. Considerando-se que o valor c da transformação depende principalmente do número k de fatores (ou $k - r$ no caso de experimento fracionado), do número m de replicações e dos coeficientes do modelo da variância, procura-se obter uma equação para determinar o valor de c em função destes valores.

O número de fatores e o número de replicações são características do projeto experimental, mas os coeficientes do modelo da variância são funções do processo em estudo e, portanto, desconhecidos. Além disso, como foi visto na seção anterior, o valor ideal de c não é o mesmo para a estimação de todos os parâmetros. Neste contexto, focaliza-se o estudo no parâmetro de maior magnitude, que normalmente é o de maior interesse.

Um projeto experimental foi desenvolvido para se obter uma equação de segundo grau que relacione o valor ótimo de c com os fatores k , m e θ_* , onde θ_* é o parâmetro de maior efeito de dispersão. Os valores codificados de k , m e θ_* foram denominados de A, D e F, respectivamente. Optou-se por usar um projeto experimental do tipo Box-Behnken (ver Myers e Montgomery, 1995, p. 318) devido os fatores k e m serem discretos e, assim, dificultar a utilização de um projeto central composto (CCD), o qual é mais comum na construção de superfícies de resposta de segunda ordem. O quadro 3.4 apresenta os níveis dos fatores utilizados neste experimento, que será referenciado como estudo de simulação nº 4.

Quadro 3.4 - Fatores e níveis usados no estudo de simulação nº 4.

Fator	Descrição	Nível inferior (-1)	Nível médio (0)	Nível superior (+1)
A	número de fatores	$k = 3$	$k = 5$	$k = 7$
D	número de replicações	$m = 1$	$m = 3$	$m = 5$
F	parâmetro	$\theta_* = 0,3$	$\theta_* = 0,7$	$\theta_* = 1,1$

Os ensaios foram feitos em dois blocos, os quais diferem em termos do número de parâmetros dos modelos. No primeiro bloco, considerou-se as equações da média e da variância como

$$\mu_y = \beta_0 + \beta_2 x_2 \quad \text{e} \quad \sigma_y^2 = \exp\{\theta_0 + \theta_1 x_1\} \quad (3.20)$$

respectivamente. Os valores dos parâmetros na simulação foram $\beta_0 = 0$, $\beta_2 = 1$, $\theta_0 = 0$ e $\theta_1 = \theta_*$ variando conforme indicado no quadro 3.4.

No segundo bloco, aumentou-se o número de parâmetros das equações. Nos modelos das médias usaram-se mais um efeito linear e uma interação nos ensaios com $k = 3$; todos os efeitos lineares e mais cinco interações nos ensaios com $k = 5$ e todos os efeitos lineares e mais sete interações nos ensaios com $k = 7$. Os efeitos lineares foram gerados com coeficientes iguais a 1 e as interações com coeficientes iguais a 0,5. Nos modelos das variâncias acrescentou-se o termo x_2 para os ensaios com $k = 3$, os termos x_2 e $x_1 \cdot x_2$ para os ensaios com $k = 5$ e os termos x_2 , x_3 e $x_1 \cdot x_2$ para os ensaios com $k = 7$. Os termos adicionais foram gerados com coeficientes iguais a 0,3.

Em todos os ensaios, as estimações dos parâmetros foram feitas pelo método RQ com duas iterações. Para cada amostra simulada, a técnica dos mínimos quadrados foi aplicada sobre os resultados da transformação $\log(\hat{e}_i^2)$ e, também, sobre os resultados de $\log(\hat{e}_i^2 + c\bar{e}^2)$, com c variando de 0,002 a 0,486, segundo uma progressão geométrica de razão 3. Para cada valor de c , comparou-se a eficiência da estimação com os dois tipos de transformação logarítmica (padrão e modificada). Simulou-se 2.000 amostras nos ensaios em que $k < 5$ ou $m < 3$ e 1.000 amostras nos ensaios com valores maiores de k e m . Cabe observar que, em cada bloco, iniciou-se a geração dos números aleatórios com a mesma semente, o que reduz a influência do erro experimental na comparação dos valores de c . O programa computacional encontra-se no apêndice.

Aos valores de eficiência relativa de cada ensaio, ajustou-se um polinômio de terceiro grau e buscou-se o valor de c que levava a função ao seu valor máximo. Nos ensaios em que o valor máximo de c ficou numa das extremidades do intervalo $[0,002, 0,486]$, ampliou-se o intervalo de investigação com simulações adicionais. A tabela 3.9 apresenta os níveis ensaiados dos fatores A, D e F e o valor máximo de c

em cada bloco. No ponto (0, 0, 0) de cada bloco, gerou-se duas seqüências de amostras simuladas, alterando-se a semente de geração dos números pseudo aleatórios.

Tabela 3.9 - Projeto experimental da simulação nº 4 e os valores de c que maximizam a eficiência da transformação logarítmica modificada (y_i representa o valor de c avaliado como ideal no bloco i , $i = 1, 2$).

#	A	D	F	y_1	y_2	#	A	D	F	y_1	y_2
1	-1	-1	0	0,555	0,770	8	1	0	1	0,002	0,002
2	-1	1	0	0,067	0,059	9	0	-1	-1	0,572	0,901
3	1	-1	0	0,005	0,005	10	0	-1	1	0,006	0,006
4	1	1	0	0,005	0,005	11	0	1	-1	0,088	0,098
5	-1	0	-1	1,120	1,102	12	0	1	1	0,005	0,004
6	-1	0	1	0,024	0,021	13	0	0	0	0,015	0,014
7	1	0	-1	0,034	0,042	14	0	0	0	0,013	0,016

Com os dados da tabela 3.9, construiu-se um modelo polinomial de segunda ordem, relacionando-se o logaritmo do valor ótimo de c com os fatores A, D e F, controlando-se as variações entre os dois blocos. O primeiro teste foi realizado para verificar se há diferença entre os blocos. Obteve-se $F_0(1, 17) = 0,43$ (nível descritivo de significância igual a 0,519), o que indica um efeito de bloco não significativo. Em função deste resultado, utilizaram-se os blocos como se fossem replicações e ajustou-se uma função quadrática. Para verificar a possibilidade de se considerar uma função mais simples, testaram-se com o apoio do procedimento *rsreg* do SAS, a contribuição adicional dos coeficientes dos termos lineares, quadráticos puros e de interação, cujos resultados são apresentados na tabela 3.10.

Tabela 3.10 - Análise de variância do ajuste de um polinômio de segundo grau para prever o valor ótimo de c , após transformação logarítmica.

Fonte de variação	gl	SQ	QM	F_0	p_F	R^2
Regressão quadrática completa	9	113,577	12,620	220,2	0,0000	0,991
termos lineares	3	106,308	35,436	618,4	0,0000	0,928
termos quadráticos	3	2,609	0,870	15,18	0,0000	0,023
termos de interação	3	4,660	1,553	27,11	0,0000	0,041
Resíduo (erro da regressão)	18	1,031	0,057			
falta de ajuste	3	0,782	0,261	15,66	0,0001	
erro puro (nas replicações)	15	0,250	0,017			

Siglas: gl (graus de liberdade), SQ (soma de quadrados), QM (quadrado médio) F_0 (estatística do teste), p_F (nível descritivo de significância) e R^2 (coeficiente de determinação).

As estatísticas da tabela 3.10 indicam que uma função de segundo grau ajusta-se muito bem aos logaritmos das respostas ($R^2 = 0,991$). Embora os termos lineares são os principais responsáveis por esta qualidade do ajustamento, os termos quadráticos e os termos de interação também se mostram significativos na construção do modelo. As duas últimas linhas da tabela 3.10 apontam que o teste de falta de ajuste é significativo. Contudo, tal significância deve-se muito mais à pequena contribuição do erro puro do que a uma baixa contribuição do modelo de regressão. A figura 3.12 apresenta uma análise dos resíduos do modelo quadrático completo.

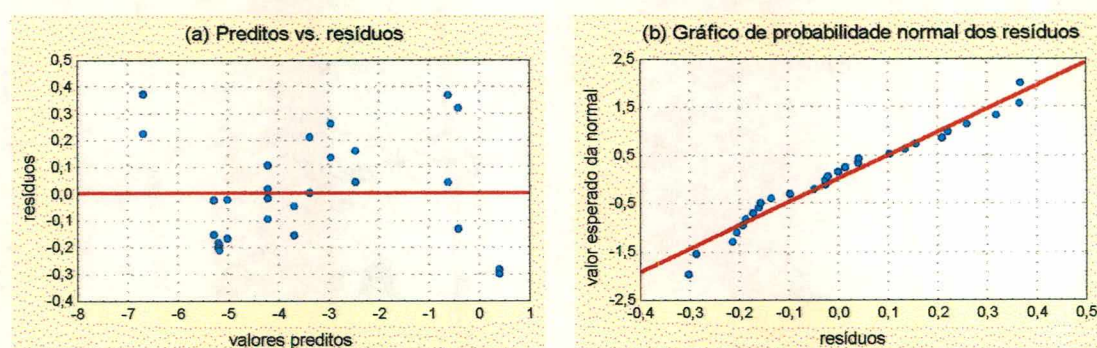


Fig. 3.12 - Análise dos resíduos do ajuste do modelo quadrático completo.

O gráfico (a) da figura 3.12 sugere que algum dos pontos mais extremos podem ter um padrão um pouco diferente dos demais. De fato, a região experimental foi bastante ampla, dificultando-se o ajuste de uma função simples, mesmo com a transformação logarítmica. Por outro lado, as estatísticas da tabela 3.10 indicam que a função quadrática é uma aproximação bastante razoável da verdadeira relação entre as variáveis em estudo. A função estimada para prever o valor ideal de c é dada por

$$\hat{c} = \exp\{-4,22 - 1,70A - 0,58D - 1,85F + 0,34A^2 + 0,39D^2 + 0,54F^2 + 0,59AD + 0,20AF + 0,45DF\}$$

(0,12) (0,06) (0,06) (0,06) (0,09) (0,09) (0,09) (0,08) (0,08) (0,08)

(3.21)

onde: $A = \frac{k-5}{2}$, sendo k o número de fatores (já descontados eventuais fracionamentos);

$D = \frac{m-3}{2}$, sendo m o número de replicações;

$F = \frac{\theta_* - 0,7}{0,4}$, sendo θ_* o coeficiente de maior magnitude do modelo da variância e

os números entre parêntesis são os erros padrão das estimativas.

Observa-se que os valores de A e D podem ser calculados antes de se proceder qualquer análise com os dados experimentais, pois dependem apenas de características do projeto experimental adotado. Por outro lado, o valor de F depende de parâmetros desconhecidos, necessitando, portanto, de uma estimativa preliminar.

Realizaram-se alguns ensaios adicionais para verificar a adequação da equação (3.21). Considerou-se um experimento do tipo 2^4 não replicado, com média e variância descritas pelas equações

$$\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \text{ e} \quad (3.22)$$

$$\sigma_y^2 = \exp\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3\} \quad (3.23)$$

com parâmetros $\beta_0 = \theta_0 = 0$, $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$, $\theta_1 = 0,896$, $\theta_2 = 0,693$ e $\theta_3 = 0,347$. A estimação dos parâmetros foi feita pelo método RQ com duas iterações.

Outro experimento usado para teste foi um 2^3 com 4 replicações, com média e variância descritas pelas equações

$$\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \text{ e} \quad (3.24)$$

$$\sigma_y^2 = \exp\{\theta_0 + \theta_1 x_1 + \theta_2 x_2\} \quad (3.25)$$

com parâmetros $\beta_0 = \theta_0 = 0$, $\beta_1 = \beta_2 = \beta_3 = 1$, $\beta_{12} = \beta_{13} = \beta_{23} = 0,5$, $\theta_1 = 1,1$, $\theta_2 = 0,5$. A estimação foi feita pelo método S2.

Em ambos os experimentos, as medidas de eficiência foram calculadas com base em 2.000 amostras simuladas de cada situação.

No experimento 2^4 , tem-se para o parâmetro θ_1 , $A = -0,5$, $D = -1$ e $F = 0,49$, levando a equação (3.21) predizer o valor $\hat{c} = 0,047$. Embora a equação (3.21) foi desenvolvida para a estimação do parâmetro de maior magnitude, observou-se também a sua aplicação para o parâmetro θ_2 , onde $F = -0,02$, fornecendo $\hat{c} = 0,14$. No experimento 2^3 , para o parâmetro θ_1 , tem-se $A = -1$, $D = 0,5$ e $F = 1$ levando ao valor predito $\hat{c} = 0,019$. Para o parâmetro θ_2 , $F = -0,5$, produzindo $\hat{c} = 0,20$. A figura 3.13 mostra as curvas de eficiência relativa da transformação proposta em função de c e as setas indicam os valores preditos pela equação (3.21).

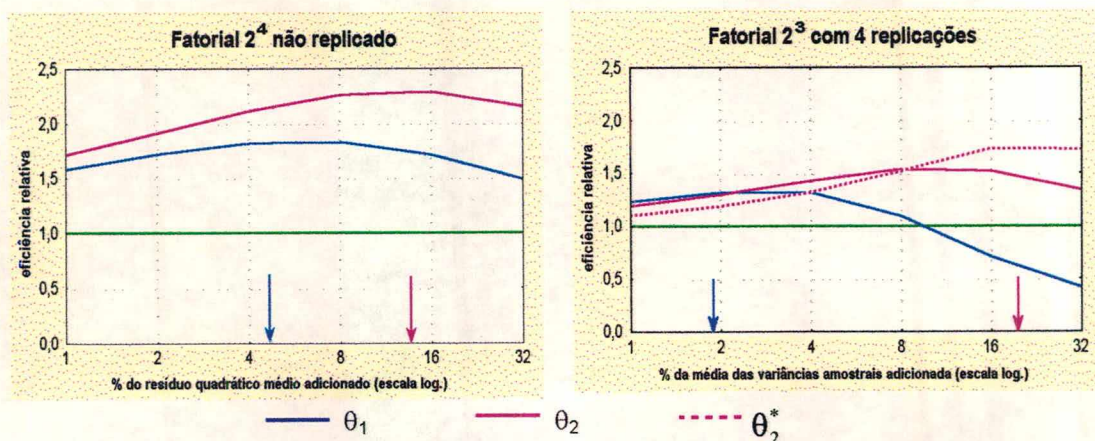


Fig. 3.13 -Eficiência da transformação logarítmica modificada em função de c e valores preditos de c pela equação (3.21).

Observa-se, na figura 3.13, que a equação (3.21) fez previsões muito boas para c em termos dos parâmetros de maior magnitude ($\theta_* = \theta_1$). Para os parâmetros de magnitude menores (θ_2), a previsão continuou boa no experimento 2^4 e ficou regular no 2^3 , onde o parâmetro θ_1 tem magnitude bastante elevada. Um ensaio adicional foi realizado no experimento 2^3 , considerando o modelo da variância como

$$\sigma_y^2 = \exp\{\theta_0 + \theta_2^* x_2\} \quad (3.26)$$

simulado com $\theta_2^* = 0,5$. A curva de eficiência deste ensaio também é mostrada na figura 3.13, onde verifica-se que o valor de c , predito pela equação (3.21), é adequado.

A análise estatística feita no ajuste da equação de regressão, assim como os ensaios adicionais, mostraram a boa adequação da equação (3.21) para especificar a transformação logarítmica na equação de regressão da variância, pelo menos para a estimação do efeito de dispersão mais forte.

3.3.4 - A transformação em experimentos do tipo 3^k e CCD

Seguindo a mesma abordagem da seção anterior, realizou-se o estudo de simulação nº 5 para buscar um valor adequado de c em experimentos do tipo fatorial 3^k e do tipo central composto de segunda ordem (CCD), que são bastante comuns em problemas onde se deseja ajustar uma superfície de resposta de segunda ordem.

O quadro 3.5 apresenta os níveis dos fatores utilizados na simulação nº 5, onde optou-se por uma redução da região experimental em termos dos fatores k e m , quando comparado ao estudo de simulação nº 4. Com esta redução, procura-se evitar situações em que o valor ótimo de c seja muito pequeno, o que pode prejudicar o ajuste da equação de regressão. Além disso, não é muito comum se ter grande número de fatores em projetos experimentais dos tipos 3^k e CCD completos.

Quadro 3.5 - Fatores e níveis usados no estudo de simulação nº 5.

Fator	Descrição	Nível inferior (-1)	Nível médio (0)	Nível superior (+1)
A	número de fatores	$k = 2$	$k = 3$	$k = 4$
D	número de replicações	$m = 1$	$m = 2$	$m = 3$
F	parâmetro	$\theta_* = 0,3$	$\theta_* = 0,7$	$\theta_* = 1,1$

As amostras foram geradas com média especificada por um modelo de segunda ordem completo, sendo os coeficientes dos termos lineares iguais à 1 e os coeficientes dos termos quadráticos e de interações iguais a 0,5. A variância, para todos os casos, foi especificada como $\sigma_y^2 = \exp\{\theta_0 + \theta_1x_1 + \theta_{11}x_1^2\}$, com $\theta_0 = 0$, $\theta_{11} = 0,3$ e $\theta_1 = \theta_*$ conforme especificado no quadro 3.5. Os valores dos fatores dos experimentos simulados foram codificados da forma usual, sendo que no caso do CCD, repetiu-se observações no ponto central para garantir precisão aproximadamente uniforme.

Para cada ponto experimental ensaiado, simulou-se 2.000 amostras, com exceção dos ensaios relativos ao experimento do tipo 3^k com $k = 4$, onde o número de amostras simuladas foi 1.000. Com estas amostras, avaliou-se a eficiência da transformação logarítmica modificada para vários valores de c e buscou-se o valor de c que maximizava esta eficiência, seguindo os mesmos procedimentos do estudo de simulação nº 4 (ver seção anterior). A tabela 3.11 mostra os resultados.

Tabela 3.11 - Projeto experimental da simulação nº 5 e os valores de c que maximizam a eficiência da transformação logarítmica modificada.

#	A	D	F	3^k	CCD	#	A	D	F	3^k	CCD
1	-1	-1	0	3,865	1,550	8	1	0	1	0,005	0,011
2	-1	1	0	0,209	0,134	9	0	-1	-1	3,350	13,066
3	1	-1	0	0,019	0,193	10	0	-1	1	0,032	0,095
4	1	1	0	0,008	0,053	11	0	1	-1	0,474	1,087
5	-1	0	-1	7,374	4,063	12	0	1	1	0,010	0,019
6	-1	0	1	0,086	0,068	13	0	0	0	0,080	0,148
7	1	0	-1	0,136	1,158	14	0	0	0	0,081	0,138

Para cada um dos tipos de experimento, ajustou-se equações de regressão para prever um valor adequado para c . No experimento fatorial 3^k , a equação ajustada apresentou $R^2 = 0,993$ e é dada por

$$\hat{c} = \exp\{-2,52 - 1,92A - 0,86D - 2,04F + 0,12A^2 + 0,16D^2 + 0,46F^2 + 0,52AD + 0,28AF + 0,19DF\}$$

(0,24) (0,12) (0,12) (0,12) (0,19) (0,19) (0,19) (0,17) (0,17) (0,17)

(3.27)

E a equação ajustada para o experimento CCD apresentou $R^2 = 0,9996$ e é dada por

$$\hat{c} = \exp\{-1,94 - 0,76A - 0,98D - 2,21F - 0,05A^2 + 0,45D^2 + 0,58F^2 + 0,29AD - 0,14AF + 0,22DF\}$$

(0,05) (0,03) (0,03) (0,03) (0,04) (0,04) (0,04) (0,04) (0,04) (0,04)

(3.28)

onde: $A = k - 3$, sendo k o número de fatores;¹
 $D = m - 2$, sendo m o número de replicações;
 $F = \frac{\theta_* - 0,7}{0,4}$, sendo θ_* o coeficiente de maior magnitude do modelo da variância e
os números entre parêntesis são os erros padrão das estimativas.

As equações (3.27) e (3.28) podem ser usadas para especificar a transformação logarítmica modificada em experimentos do tipo 3^k e CCD, desde que se tenha uma estimativa inicial do parâmetro de maior efeito de dispersão, referenciado por θ_* .

3.3.5 - Estudo do vício das estimativas

Realizou-se uma análise complementar para estudar o vício das estimativas dos parâmetros do modelo da variância. O estudo utilizou-se de um planejamento fatorial 2^4 e um central composto de segunda ordem com 2 fatores, que foram observados em 4 pontos fatoriais, 4 axiais e 5 centrais. Em ambos os casos, consideraram-se 2 replicações.

Os modelos de simulação para o experimento 2^4 foram:

$$\mu_y = x_1 + x_2 + x_3 + x_4 + 0,5(x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4) \quad (3.29)$$

¹ No experimento do tipo 3^k , pode-se considerar k como o número de fatores menos r (número de fracionamentos).

$$\sigma_y^2 = \exp\{(0,693)x_1 + (0,549)x_2 + (0,347)x_3\}$$

(3.30)

e para o experimento CCD foram:

$$\mu_y = x_1 + x_2 + 0,5(x_1^2 + x_2^2 + x_1x_2)$$

(3.31)

$$\sigma_y^2 = \exp\{(0,693)x_1 + (0,549)x_2 + (0,347)x_1^2\}$$

(3.32)

com os fatores codificados em ± 1 nos pontos fatoriais, 0 nos pontos centrais e $\pm 1,414$ nos pontos axiais. A estimato dos parmetros dos modelos das varincias foi feita pelo mtodo S2, sendo que no ponto central do experimento CCD, calcularam-se varincias amostrais a cada dupla de observaes. Com base em 5.000 amostras simuladas de cada experimento, calcularam-se os coeficientes de variao e os vcios relativos das estimativas. A tabela 3.12 apresenta estes resultados em funo dos valores de c .

Tabela 3.12 - Coeficiente de variao e vcio relativo dos estimadores com a transformao proposta.

projeto experimental	c	coeficiente de variao			vcio relativo (%)		
		θ_1	θ_2	$\theta_3 (\theta_{11})$	θ_1	θ_2	$\theta_3 (\theta_{11})$
2 ⁴	0,00	0,79	1,00	1,53	0,01	0,02	0,02
	0,01	0,67	0,86	1,31	-0,14	-0,14	-0,14
	0,04	0,62	0,81	1,24	-0,26	-0,26	-0,26
	0,16	0,57	0,77	1,19	-0,44	-0,44	-0,43
	0,64	0,54	0,73	1,16	-0,65	-0,65	-0,64
	2,56	0,51	0,72	1,17	-0,83	-0,83	-0,82
CCD	0,00	1,15	1,43	2,50	-0,01	0,01	-0,03
	0,01	1,00	1,22	2,10	-0,13	-0,14	-0,13
	0,04	0,94	1,15	1,93	-0,23	-0,25	-0,21
	0,16	0,88	1,09	1,75	-0,39	-0,43	-0,35
	0,64	0,83	1,05	1,58	-0,60	-0,64	-0,54
	2,56	0,80	1,06	1,48	-0,79	-0,83	-0,76

A tabela 3.12 mostra que o coeficiente de variao de todos os coeficientes diminui com o aumento de c , ou seja, nas condies experimentais estudadas tem-se menor variabilidade relativa com valores altos de c . Quanto ao vcio, observa-se, na primeira linha de ambos os experimentos, que sob a transformao logartmica padro, o vcio pode ser negligenciado. Por outro lado, ao estabelecer a equao por $\log(S_i^2 + c \overline{S}^2)$, com $c > 0$, as estimativas passam a ser viciadas e o vcio aumenta de magnitude na medida em que se aumenta c .

Uma caracterstica particularmente interessante no experimento fatorial 2⁴  que, para um dado $c > 0$, o vcio relativo  negativo e aproximadamente constante ao

longo dos parâmetros. No experimento CCD, o vício relativo altera-se ligeiramente em função dos parâmetros. Isto pode ser devido ao fato dos efeitos quadráticos não serem estimados independentemente dos lineares, enquanto que num projeto fatorial do tipo 2^k todos os efeitos são ortogonais.

Existem algumas situações em que o vício, quando provocado por um fator constante, não prejudica o resultado final da análise. Por exemplo, quando se busca a combinação de níveis dos fatores que minimizam a variância, um fator constante não altera o resultado final. Outra situação é o processo de identificação dos efeitos não negligenciáveis através do gráfico normal dos efeitos ou do gráfico de Pareto, pois, nestes gráficos, analisa-se a magnitude dos efeitos de forma relativa e, assim, um vício relativo constante não prejudica a análise. Logo, nestas situações, pode-se usar valores maiores para c do que os preditos pelas equações das seções anteriores.

3.4 - AGREGAÇÃO DAS VARIÂNCIAS AMOSTRAIS COM OS RESÍDUOS QUADRÁTICOS MÉDIOS

Nas seções anteriores, consideraram-se os métodos S2 e RQ como formas alternativas para se estimarem os parâmetros do modelo da variância. Nesta seção, apresenta-se uma proposta de se agregarem os dois métodos, ou seja, procura-se uma forma de combinar, em cada ponto experimental, as informações da variância amostral e do resíduo quadrático médio.

3.4.1 - O método de agregação

A proposta de combinar os métodos S2 e RQ baseia-se no fato de que quando o modelo da média estiver corretamente especificado, o uso dos resíduos quadráticos como informação primária para se estimar o modelo da variância deve produzir melhores resultados, quando comparado à utilização das variâncias amostrais. Por outro lado, se o modelo da média estiver inadequado, a utilização dos resíduos quadráticos deve produzir maus resultados.

Desde que o experimento tenha replicações genuínas em todos os pontos experimentais, propõe-se considerar os resultados da seguinte estatística como base para a estimação do modelo de σ_y^2 :

$$v_i = \gamma \hat{e}_i^2 + (1-\gamma)S_i^2, \quad i = 1, 2, \dots, M \quad (3.33)$$

onde: S_i^2 é a variância amostral no i -ésimo ponto experimental;

\hat{e}_i^2 é o residuo quadrático médio no i -ésimo ponto experimental e

γ é um grau de pertinência da adequação do modelo de regressão da média.

Propõe-se que o valor de γ seja calculado a partir dos próprios dados, considerando o teste estatístico de falta de ajuste do modelo de regressão da média. Seja p_F o nível descritivo de significância produzido pelo valor da estatística F_0 do teste de falta de ajuste, isto é,

$$p_F = P\{F_{(M-p, n-M)} \geq F_0\} \quad (3.34)$$

onde: M é o número de pontos experimentais ensaiados;

p é o número de parâmetros do modelo de regressão da média;

$n = Mm$ é o número total de observações;

$F_{(M-p, n-M)}$ é uma variável aleatória com distribuição F de $(M - p)$ graus de liberdade no numerador e $(n - M)$ graus de liberdade no denominador e

F_0 é a estatística do teste de falta de ajuste, expressão (2.12).

Sugere-se que γ seja calculado por

$$\gamma = \begin{cases} 1 & \text{se } p_F \geq b \\ (b - p_F)/(b - a) & \text{se } a \leq p_F < b \\ 0 & \text{se } p_F \leq a \end{cases} \quad (3.35)$$

onde $0 < a \leq b < 1$.

Quando $a = b$, tem-se um estimador do tipo *pré-teste*. Esta categoria de estimador é bastante conhecida na literatura de estatística e de econometria (ver, por exemplo, Judge et al., 1988, p. 362). No presente contexto, utiliza-se uma abordagem mais difusa, conforme definida em Klir e Yuan (1995). A escolha da abordagem difusa justifica-se pelo fato de que a definição de *modelo adequado* pode ser um tanto quanto vaga. Além disso, o teste de falta de ajuste, que fornece a estatística básica para o cálculo de p_F , não é propriamente válido para modelos heterocedásticos (de variância não constante), como é o caso em estudo.

A tentativa de aprimorar procedimentos através da construção de uma medida resultante da combinação linear de duas outras medidas já vem sendo explorada por outros autores, em situações parecidas. Hooper (1993) propõe combinar linearmente a variância amostral e a variância predita por um modelo de regressão para ponderar as observações e, com isto, corrigir problemas de heterocedasticidade. A combinação linear proposta por Hooper baseia-se no princípio da máxima verossimilhança. Liu et al. (1995) combinam a precisão das predições de um modelo (obtidas objetivamente dos dados) com o grau de crença do modelo (obtida subjetivamente) para avaliar o nível de confiança (ou de crença) de resultados preditos.

Como as estatísticas v_i ($i = 1, 2, \dots, M$) dependem do modelo de regressão da média, sugere-se o seguinte procedimento iterativo:

- a) calculam-se as variâncias em cada ponto experimental, S_i^2 ($i = 1, 2, \dots, M$);
- b) estimam-se os parâmetros do modelo da média por mínimos quadrados ordinários (MQO);
- c) calcula-se o grau de pertinência γ da adequação do modelo da média;
- d) calculam-se, em cada ponto experimental, a média dos resíduos quadráticos (\hat{e}_i^2) e a estatística $v_i = \gamma \hat{e}_i^2 + (1-\gamma)S_i^2$ ($i = 1, 2, \dots, M$);
- e) estimam-se os parâmetros do modelo da variância por MQO, usando como variável dependente $\log(v_i + q)$, conforme descrito na seção 3.3;
- f) estimam-se, novamente, os parâmetros do o modelo da média, mas agora usando mínimos quadrados generalizados (MQG), atribuindo às observações os correspondentes pesos $w_{ij} = 1/\hat{\sigma}_i^2$ ($i = 1, 2, \dots, M; j = 1, 2, \dots, m$), onde $\hat{\sigma}_i^2$ são as variâncias preditas pelo modelo de σ_y^2 , obtido em (e);
- g) repetem-se os passos (c) - (f) pelo menos mais uma vez.

Cabe observar que, a partir da segunda iteração, o nível descritivo de significância do teste de falta de ajuste, p_F , usado no cálculo de γ , baseia-se num modelo em que o efeito da heterocedasticidade está sendo corrigido e, portanto, pode ser considerado válido, pelo menos assintoticamente.

Em experimentos não replicados, o método ainda pode ser usado se houver algum fator não significativo no modelo de σ_y^2 , o qual pode ser usado como se fosse

replicação e, assim, permitir o cálculo das variâncias amostrais. Mas, neste caso, não se tem elementos para avaliar a qualidade do modelo através do teste de falta de ajuste. Em muitas situações, contudo, o experimento é realizado com replicações no ponto central. Isto possibilita obter uma estimativa da variância do processo e, assim, efetuar um teste de falta de ajuste (ver Khuri e Cornell, 1996).

3.4.2 - Estudos de simulação

Para se verificar o desempenho do procedimento proposto, realizou-se estudos de simulação do tipo Monte Carlo sobre um experimento fatorial 2^4 completo, com 2 replicações. As respostas foram geradas com distribuição normal de média e variância dadas pelas seguintes equações:

$$\mu_y = x_1 + x_2 + x_3 + x_4 + \delta(x_1x_2 + x_2x_3 + x_3x_4) \quad (3.36)$$

$$\sigma_y^2 = \exp\{(0,693)x_1 + (0,549)x_2 + (0,347)x_3\} \quad (3.37)$$

onde x_1, \dots, x_4 recebem os valores -1 e $+1$, conforme um experimento fatorial 2^4 e δ é o coeficiente das interações do modelo.

No processo de estimação, os modelos da média e da variância foram especificados da seguinte maneira:

$$\mu_y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4; \quad (3.38)$$

$$\sigma_y^2 = \exp\{\theta_0 + \theta_1x_1 + \theta_2x_2 + \theta_3x_3\} \quad (3.39)$$

Observando-se a equação (3.38), verifica-se que, na fase de estimação, não se estão incluindo as interações geradas pela equação (3.36). Desta forma, o modelo da média é especificado erroneamente e o valor de δ indica o grau de violação na especificação do modelo.

Consideraram-se os seguintes valores de δ : 0, 0,1, 0,2, ..., 0,7. Com δ acima de 0,7 verificou-se que, ao nível de significância de 5%, o teste de falta de ajuste, após a correção da heterocedasticidade, detecta erro de especificação do modelo com probabilidade superior a 90%. Para cada valor de δ , fizeram-se 1.000 realizações simuladas do experimento.

O processo de estimação foi realizado pelos métodos S2, RQ e pela agregação de ambos (S2RQ). Em todos os casos, utilizou-se a transformação logarítmica modificada. No método S2, a estimação procedeu-se com a utilização de MQO,

enquanto nos métodos RQ e S2RQ, utilizou-se MQG com duas iterações. No método S2RQ, o cálculo de γ foi feito de duas maneiras. Na primeira, usou-se $a = b = 0,05$, caracterizando um estimador do tipo pré-teste, ou seja, quando os dados indicam a validade do modelo, utilizam-se os resíduos quadráticos médios, em caso contrário, as variâncias amostrais. Outra forma do cálculo de γ foi feito com $a = 0,01$ e $b = 0,50$, onde a divisão entre os dois procedimentos deixa de ser tão brusca, pois em quase 50% dos casos utilizam-se médias ponderadas entre as variâncias amostrais e os resíduos quadráticos médios. O programa computacional encontra-se no apêndice.

A figura 3.14 mostra o erro quadrático médio da estimativa de θ_1 em função do nível de violação do modelo (δ), considerando os três métodos (S2, RQ e S2RQ), e, no caso do método S2RQ, as duas formas de avaliação de γ . Na transformação logarítmica, adicionou-se no argumento da função a quantidade $q = (0,07)\bar{S}^2$, conforme discutido na seção 3.3.

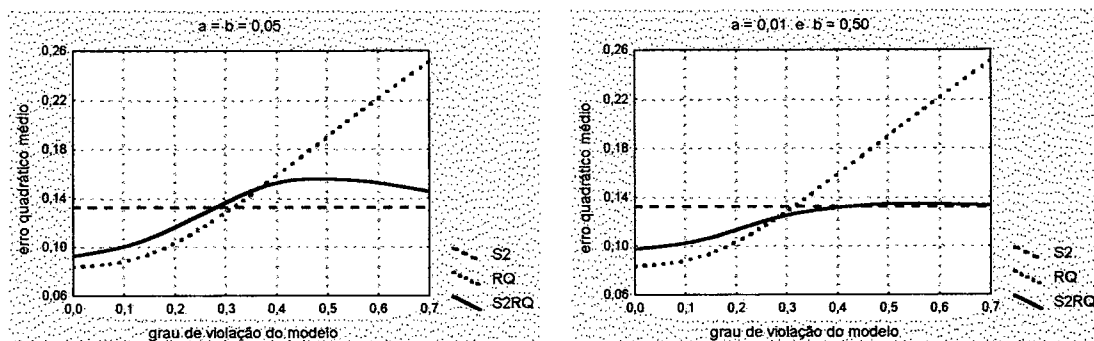


Fig. 3.14 - Comparação dos métodos S2, RQ e S2RQ em função de δ . Modelos especificados pela transformação logarítmica modificada.

A figura 3.14 mostra que o método RQ tem o menor erro quadrático médio quando o modelo está corretamente especificado, mas apresenta-se muito ruim quando o grau de violação do modelo, δ , é muito grande. Por outro lado, observa-se que a agregação dos dois métodos (S2RQ) comporta-se de maneira interessante, pois quando o modelo da média está corretamente especificado, o procedimento S2RQ produz estimativas quase tão precisas quanto às produzidas pelo método RQ e, em caso de má especificação, o método não se degenera como o RQ. Verifica-se, também, que, quando o grau de violação do modelo aumenta, a agregação com $a = 0,01$ e $b = 0,50$, que caracteriza de forma mais suave a diferença entre os conceitos de

modelo adequado e modelo inadequado, comporta-se melhor do que a abordagem de pré-teste ($a = b = 0,05$).

Uma análise complementar foi feita com outros tipos de experimentos, considerando-se o cômputo de γ com $a = 0,01$ e $b = 0,50$ e transformação logarítmica modificada. Em todos os casos, considerou-se δ aumentando em 0,1 unidade até que, ao nível de significância de 5%, o teste de falta de ajuste, após a correção da heterogeneidade, detecta erro de especificação do modelo com probabilidade superior a 90%. Os experimentos utilizados foram:

- a) O mesmo experimento descrito anteriormente, mas com 4 replicações.
- b) O mesmo tipo de experimento, com 2 replicações, mas aumentando o coeficiente focal, θ_1 , de 0,693 para 1,1.

- c) Experimento do tipo 2⁵, com o modelo da média gerado por

$$\mu_y = x_1 + x_2 + x_3 + x_4 + x_5 + \delta(x_1x_2 + x_2x_3 + x_3x_4) \quad (3.40)$$

e especificado em termos dos fatores principais. E o modelo da variância como descrito na expressão (3.37).

- d) Experimento central composto de segunda ordem com 2 fatores e 2 replicações. Modelo da média gerado por

$$\mu_y = x_1 + x_2 + \delta(x_1^2 + x_2^2 + x_1x_2) \quad (3.41)$$

e especificado somente com os termos lineares. Modelo da variância gerado por

$$\sigma_y^2 = \exp\{(0,693)x_1 + (0,549)x_2 + (0,347)x_1^2\} \quad (3.42)$$

Em todos os experimentos, utilizou-se a transformação logarítmica modificada. Nesta transformação, os valores usados para c foram 0,03, 0,01, 0,02 e 0,30 para os casos (a), (b), (c) e (d), respectivamente. Os três primeiros valores foram obtidos pela equação (3.21) e o último pela equação (3.28).

Para cada um dos experimentos, a figura 3.15 mostra os erros quadráticos médios das estimativas de θ_1 , as quais foram obtidas pelos três métodos em discussão (S2, RQ e S2RQ) e em função do nível de violação do modelo da média (δ). Os resultados baseiam-se em 1.000 realizações simuladas para cada um dos quatro experimentos e cada valor de δ . As escalas dos gráficos foram adaptadas de acordo com as características de cada experimento.

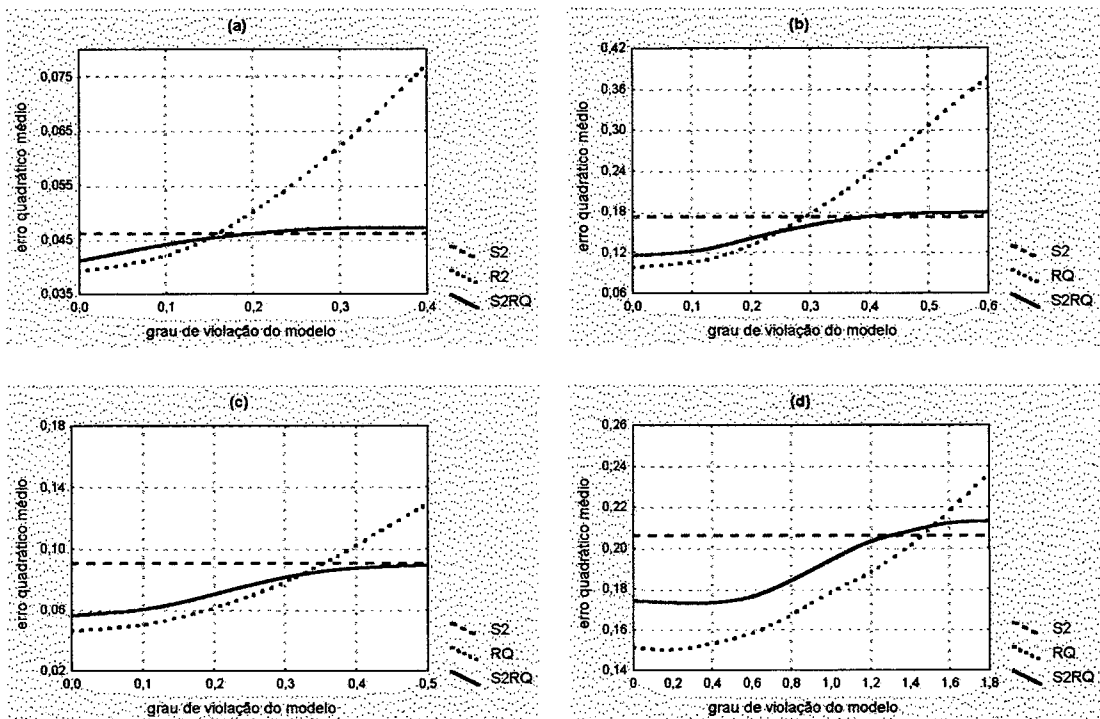


Fig. 3.15 - Comparação entre os métodos S2, RQ e S2RQ em quatro projetos experimentais.

Observa-se, na figura 3.15, que o comportamento dos métodos apresentam-se de forma bastante parecida nos quatro projetos experimentais. Nos experimentos (a) e (b), o método RQ torna-se extremamente ruim com modelos mal especificados, mesmo antes do teste de falta de ajuste detectar o problema com alto grau de probabilidade. Já nos experimentos (c) e (d), o método RQ só se degenera quando o poder do teste de falta de ajuste apresenta-se num nível elevado. Por outro lado, em todos os experimentos estudados, o método S2RQ apresenta eficiência quase tão boa quanto ao método RQ quando o modelo da resposta é corretamente especificado e, ao mesmo tempo, robustez em relação à má especificação do modelo.

ABORDAGEM MULTIVARIADA

Um produto ou processo costuma ser avaliado por múltiplas características de qualidade (y_1, y_2, \dots, y_L). Por exemplo, a qualidade de um material pode ser vista em termos de sua resistência, elasticidade, porosidade, etc. Estas características normalmente são correlacionadas e podem ser influenciadas pelos fatores do processo (x_1, x_2, \dots, x_k) por diferentes formas funcionais.

Na seção 2.3.2, estudou-se a abordagem dos sistemas de equações aparentemente não relacionadas (SUR), que permite estimar as equações de regressão das várias respostas em conjunto. Neste capítulo, procura-se incluir as equações de regressão das variâncias, formando-se dois sistemas de equações: o das médias e o das variâncias.

4.1 - O MODELO GERAL

Sejam: $\ell = 1, 2, \dots, L$ o índice que referencia as diferentes equações do sistema;

$i = 1, 2, \dots, M$ o índice que indica os pontos experimentais ensaiados e

$j = 1, 2, \dots, m$ o índice das replicações, sendo $n = Mm$ o número total de observações.

O sistema de equações das respostas e o sistema de equações das variâncias podem ser escritos, respectivamente, como

$$y_{\ell ij} = x_{\ell ij}' \beta_{\ell} + \sigma_{\ell i} e_{\ell ij} \quad (4.1)$$

$$g(\sigma_{\ell i}^2) = z_{\ell ij}' \theta_{\ell} \quad (\ell = 1, 2, \dots, L) \quad (4.2)$$

onde: $y_{\ell ij}$ são os dados da ℓ -ésima resposta;

$x_{\ell ij}$ são os vetores formados pelas linhas da matriz de planejamento X_{ℓ} , referentes à ℓ -ésima equação do sistema de regressão das respostas, com dimensão $n \times p_{\ell}$;

- β_ℓ é o vetor de parâmetros da ℓ -ésima equação do sistema de regressão das respostas, com dimensão $p_\ell \times 1$;
- $\sigma_{\ell i}^2$ são as variâncias nos M pontos experimentais, referentes às L respostas;
- g é alguma função que permita representar as variâncias através de modelos lineares;
- $z_{\ell i}$ são as linhas da matriz de planejamento Z_ℓ , referentes à ℓ -ésima equação do sistema de regressão das variâncias, com dimensão $M \times q_\ell$;
- θ_ℓ é o vetor de parâmetros da ℓ -ésima equação do sistema das variâncias, com dimensão $q_\ell \times 1$ e
- $e_{\ell ij}$ são os erros aleatórios, os quais supostamente têm, para cada equação, valor esperado nulo, variância unitária e, preferencialmente, distribuição normal;

O sistema de equações (4.1), quando visto em termos de valor esperado, será referenciado como sistema de equações das médias do processo.

4.1 - O SISTEMA DE EQUAÇÕES DAS MÉDIAS

Expressando (4.1) sob um modelo ampliado do tipo SUR, tem-se

$$y = X\beta + \varepsilon \quad (4.3)$$

onde:

$y = (y_{11}, y_{12}, \dots, y_{1n}, y_{21}, y_{22}, \dots, y_{2n}, \dots, y_{L1}, y_{L2}, \dots, y_{Ln})'$ é o vetor composto pelas nL observações, agregadas sequencialmente;

$$X = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & X_L \end{bmatrix}$$

é a matriz de planejamento ampliada, de dimensão $nL \times P$ (sendo $P = \sum_{\ell=1}^L p_\ell$), formada

pelas matrizes de planejamento X_ℓ , com dimensões $n \times p_\ell$ ($\ell = 1, 2, \dots, L$) e matrizes nulas, 0, de dimensões compatíveis;

$\beta = (\beta_1', \beta_2', \dots, \beta_L')' = (\beta_{11}, \dots, \beta_{1p_1}, \beta_{21}, \dots, \beta_{2p_2}, \beta_{L1}, \dots, \beta_{Lp_L})'$ é o vetor de parâmetros com P elementos, isto é, os parâmetros das L equações agregados sequencialmente num único vetor e

$\varepsilon = (\sigma_1 \cdot e_{11}, \dots, \sigma_1 \cdot e_{1n}, \sigma_2 \cdot e_{21}, \dots, \sigma_2 \cdot e_{2n}, \dots, \sigma_L \cdot e_{L1}, \dots, \sigma_L \cdot e_{Ln})'$ é o vetor de erros aleatórios, também ampliado.

4.2.1 - O modelo SUR com variâncias desiguais

Admitindo-se que as variâncias possam mudar ao longo das diferentes condições experimentais (heterocedasticidade), a estrutura da matriz de variâncias e covariâncias, Σ^* , torna-se um pouco mais complexa do que a apresentada na seção 2.3.2.

Sejam $\ell, k = 1, 2, \dots, L$; $i = 1, 2, \dots, M$ e $j = 1, 2, \dots, n$, sendo $n = Mm$. Considere, também, as seguintes representações:

$\sigma_{\ell\ell}^{(j)}$ é a variância de y_ℓ na j -ésima observação e

$\sigma_{\ell k}^{(j)}$ é a covariância entre y_ℓ e y_k na j -ésima observação.

Quando o experimento é realizado com m replicações, tem-se $\sigma_{\ell k}^{(i)} = \sigma_{\ell k}^{(i+1)} = \dots = \sigma_{\ell k}^{(i+m-1)}$.

Com esta notação, a matriz Σ^* pode ser escrita da seguinte forma:

$$\Sigma^* = \begin{bmatrix} \sigma_{11}^{(1)} & 0 & \dots & 0 & \dots & \sigma_{1L}^{(1)} & 0 & \dots & 0 \\ 0 & \sigma_{11}^{(2)} & \dots & 0 & \dots & 0 & \sigma_{1L}^{(2)} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{11}^{(n)} & \dots & 0 & 0 & \dots & \sigma_{1L}^{(n)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{1L}^{(1)} & 0 & \dots & 0 & \dots & \sigma_{LL}^{(1)} & 0 & \dots & 0 \\ 0 & \sigma_{1L}^{(2)} & \dots & 0 & \dots & 0 & \sigma_{LL}^{(2)} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{1L}^{(n)} & \dots & 0 & 0 & \dots & \sigma_{LL}^{(n)} \end{bmatrix} \quad (4.4)$$

Para se obter o estimador $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2', \dots, \hat{\beta}_L')'$ dos coeficientes das equações do sistema, conforme apresentado na seção 2.3.2, precisa-se estimar os elementos $\sigma_{\ell k}^{(j)}$ de Σ^* . A maior dificuldade é encontrar bons estimadores para todas as $\sigma_{\ell k}^{(j)}$ ($k, \ell = 1, 2, \dots, L$; $j = 1, 2, \dots, n$). Srivastava e Giles (1987, cap. 8) sugerem realizar a estimação das variâncias e covariâncias em subgrupos homocedásticos da amostra. Eles também apontam para a possibilidade de se obterem estimativas de máxima

verossimilhança usando-se o método descrito em Magnus (1978), que é suficientemente genérico para abrigar o sistema SUR com variâncias desiguais. As sugestões de Srivastava e Giles, contudo, voltam-se para estudos econométricos, onde é comum se ter grande número de observações. Isto porque as boas propriedades dos métodos sugeridos são assintóticas, ou seja, voltadas para quando n cresce infinitamente. Já em estudos experimentais, normalmente não se tem número suficiente de observações para se estimarem adequadamente todas as $\sigma_{\epsilon k}^{(j)}$.

Uma outra abordagem é apresentada em Mandy e Martins-Filho (1993), que sugerem construir modelos para todas as variâncias e covariâncias. Por esta abordagem, o sistema de equações das variâncias (4.2) é ampliado para contemplar também as equações das covariâncias. Com as predições destes modelos, pode-se ter uma estimativa para a matriz Σ^* e efetuar a estimação dos coeficientes do sistema. Cabe observar, porém, que este método pode levar a uma análise muito complexa se o sistema for composto de muitas equações.

4.2.2 - A suposição de correlações constantes

Neste trabalho, propõe-se inserir uma suposição adicional, que permita reduzir o número de parâmetros da matriz de variâncias e covariâncias. A suposição é que as correlações entre os termos de erro não se alteram ao longo das diferentes condições experimentais. Ou seja, considera-se que as correlações são características intrínsecas das variáveis do sistema e não são produzidas ou alteradas pelas diferentes condições experimentais.

A suposição em discussão aparece em trabalhos com problemas parecidos. Vaccaro et al. (1996), no desenvolvimento de um algoritmo de regressão múltipla multivariada, também admitiu implicitamente uma hipótese similar. Zeger e Liang (1986) e Chaganty (1996) usam uma extensão de modelos lineares generalizados para a análise de dados longitudinais, onde adotam uma suposição parecida com a que se está descrevendo. Eles consideram um vetor y de medidas repetidas e um vetor x de covariáveis, onde a média de y está associada às covariáveis por um vetor de parâmetros β e a variância de y está associada à média por um parâmetro ϕ . Por

outro lado, as correlações entre as medidas repetidas se relacionam entre si por um vetor de parâmetros α , mas não se relacionam com x ou com a média de y .

Outra suposição parecida foi feita por Hughes-Oliver et al. (1998), que se utilizaram de um modelo multivariado para buscar uniformidade num processo de fabricação de semicondutores. Conforme descrevem os autores, uma fase crítica do processo consiste no depósito de uma finíssima camada de polissilício, a qual forma uma micro bolacha de silício (*silicon wafer*). A espessura de silício deve ser tão uniforme quanto possível e estar suficientemente próxima do valor ideal (1000 *angstroms*). Um projeto experimental foi conduzido, onde observou-se medidas de espessura em 13 locais previamente estabelecidos de cada bolacha. Para obter equações de regressão para a média e a variância do processo, os autores admitiram que a média e a variância eram funções dos fatores do processo, enquanto as correlações entre as medidas de uma mesma bolacha eram funções apenas da localização espacial. A estimação foi feita pelo método da máxima verossimilhança restrita.

Em termos matemáticos, a suposição de que as correlações não se alteram ao longo das diferentes condições experimentais pode ser expressada como

$$\rho_{k\ell}^{(j)} = \rho_{k\ell} \quad (4.5)$$

onde $\rho_{k\ell}^{(j)}$ é a correlação entre os termos de erro das equações de y_k e y_ℓ ($k, \ell = 1, 2, \dots, L$) referente à j -ésima observação ($j = 1, 2, \dots, n$).

Seja $\rho = \{\rho_{k\ell}\}$ a matriz de dimensão $L \times L$, cujos os elementos são as correlações dos termos de erro de y_1, y_2, \dots, y_L . Para colocar o estimador de MQG de β em termos de ρ , considere uma matriz diagonal P , de dimensão $nL \times nL$, formada pelo inverso dos desvios padrão dos termos de erro de $y_{\ell j}$ ($\ell = 1, 2, \dots, L; j = 1, 2, \dots, n$), ou seja:

$$P = \text{diag} \left(\frac{1}{\sqrt{\sigma_{11}^{(1)}}}, \dots, \frac{1}{\sqrt{\sigma_{11}^{(n)}}}, \dots, \frac{1}{\sqrt{\sigma_{\ell\ell}^{(1)}}}, \dots, \frac{1}{\sqrt{\sigma_{\ell\ell}^{(n)}}}, \dots, \frac{1}{\sqrt{\sigma_{LL}^{(1)}}}, \dots, \frac{1}{\sqrt{\sigma_{LL}^{(n)}}} \right) \quad (4.6)$$

Pré multiplicando o modelo SUR de (4.3) por P , tem-se:

$$P y = P X \beta + P \varepsilon \quad (4.7)$$

ou

$$y_T = X_T \beta + \varepsilon_T \quad (4.8)$$

onde $y_T = P y$, $X_T = P X$ e $\varepsilon_T = P \varepsilon$.¹ Observa-se que, em termos das variáveis transformadas, a matriz de variâncias e covariâncias dos termos de erro pode ser escrita como

$$\text{Var}\{\varepsilon_T\} = \text{Var}\{P \varepsilon\} = P \Sigma^* P' \quad (4.9)$$

onde a última igualdade vem das propriedades da variância de um vetor (ver, por exemplo, Johnson e Wichern, 1992, p. 62). Efetuando-se as multiplicações matriciais, tem-se:

$$\text{Var}\{\varepsilon_T\} = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & \rho_{1L}^{(1)} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 & \rho_{1L}^{(2)} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \dots & 0 & 0 & \dots & \rho_{1L}^{(n)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{1L}^{(1)} & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \rho_{1L}^{(2)} & \dots & 0 & \dots & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \rho_{1L}^{(n)} & \dots & 0 & 0 & \dots & 1 \end{bmatrix} \quad (4.10)$$

onde $\rho_{kl}^{(j)} = \frac{\sigma_{kl}^{(j)}}{\sqrt{\sigma_{kk}^{(j)}} \sqrt{\sigma_{ll}^{(j)}}}$ é a correlação entre y_k e y_l , referente à j -ésima observação.

Nota-se que a matriz $\text{Var}\{\varepsilon_T\}$ é formada por blocos de matrizes identidade ao longo da diagonal principal e blocos de matrizes diagonais de correlações fora da diagonal principal. Considerando a suposição (4.5), a qual afirma que as correlações são iguais para todas as observações ($\rho_{kl}^{(j)} = \rho_{kl}$), pode-se escrever

¹ Esta transformação, quando aplicada em modelos heterocedásticos de regressão univariada, faz parte do método MQG. A soma de quadrados dos erros no modelo transformado corresponde à soma ponderada de quadrados no modelo original, sendo que a ponderação é dada pelo inverso das respectivas variâncias.

$$\text{Var}\{\varepsilon_T\} = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & \rho_{1L} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 & \rho_{1L} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \dots & 0 & 0 & \dots & \rho_{1L} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho_{1L} & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \rho_{1L} & \dots & 0 & \dots & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \rho_{1L} & \dots & 0 & 0 & \dots & 1 \end{bmatrix} = \rho \otimes I_n \quad (4.11)$$

onde: $\rho = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1L} \\ \rho_{21} & 1 & \dots & \rho_{2L} \\ \dots & \dots & \dots & \dots \\ \rho_{L1} & \rho_{L2} & \dots & 1 \end{bmatrix}$ é a matriz de correlações dos termos de erro e_1, e_2, \dots, e_L , relativos às equações das respostas y_1, y_2, \dots, y_L e

$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}$ é a matriz identidade de ordem n .

Observando que (4.8) é um modelo SUR com matriz de variâncias e covariâncias dos termos de erro dada por (4.11), o estimador de MQG de β pode ser escrito como²

$$\tilde{\beta} = [X_T'(\rho^{-1} \otimes I_n)X_T]^{-1} X_T'(\rho^{-1} \otimes I_n)y_T \quad (4.12)$$

O estimador $\tilde{\beta}$ torna-se viável na medida em que os elementos da matriz ρ são substituídos por estimadores. Observa-se, porém, que o número de parâmetros a estimar em (4.11) é bem menor do que em (4.4), viabilizando o processo de estimação também para experimentos com poucas replicações.

4.2.3 - Estimação dos parâmetros

Para se estimarem os elementos de ρ , precisa-se, inicialmente, aplicar o método MQG sobre cada equação do sistema das respostas. Nota-se que, com a

² Este estimador foi apresentado na seção 2.3.2 em termos de Σ e num modelo homocedástico como $\tilde{\beta} = [X'(\Sigma^{-1} \otimes I_n)X]^{-1} X'(\Sigma^{-1} \otimes I_n)y$.

aplicação do MQG, tem-se uma estimativa da matriz diagonal P e, portanto, pode-se efetuar, pelo menos aproximadamente, as transformações $y_T = P y$ e $X_T = P X$, necessárias para a aplicação da fórmula (4.12). As transformadas y_T e X_T de cada equação são obtidas implicitamente no método MQG univariado.

Com os resíduos $\hat{e}_\ell^{(j)}$ ($\ell = 1, 2, \dots, L; j = 1, 2, \dots, n$) das equações de regressão das médias, obtêm-se estimativas para as correlações $\rho_{k\ell}$ pelo método dos momentos de Pearson. Este coeficiente é discutido, por exemplo, em Barbetta (1998, p. 248) e pode ser calculado, em termos dos $\hat{e}_\ell^{(j)}$, segundo a expressão:

$$\hat{\rho}_{k\ell} = \frac{1}{n-1} \sum_{j=1}^n \left(\frac{\hat{e}_k^{(j)} - \bar{\hat{e}}_k}{s_{\hat{e}_k}} \right) \left(\frac{\hat{e}_\ell^{(j)} - \bar{\hat{e}}_\ell}{s_{\hat{e}_\ell}} \right) \quad (4.13)$$

onde: $\bar{\hat{e}}_k$ e $\bar{\hat{e}}_\ell$ são as médias aritméticas dos $\hat{e}_k^{(j)}$ e $\hat{e}_\ell^{(j)}$, respectivamente e $s_{\hat{e}_k}$ e $s_{\hat{e}_\ell}$ são os desvios padrão dos $\hat{e}_k^{(j)}$ e $\hat{e}_\ell^{(j)}$, respectivamente.

Os termos $\bar{\hat{e}}_k$ e $\bar{\hat{e}}_\ell$ são normalmente nulos e poderiam ser excluídos de (4.13). Contudo, prefere-se adotar a fórmula mais geral para também contemplar resíduos provindos das equações estimadas sob o sistema SUR, onde os resíduos deixam de ter média nula quando analisados em cada equação isoladamente.

Chamando-se de $\hat{\rho}$ a matriz $L \times L$, formada pelas correlações estimadas $\hat{\rho}_{k\ell}$ ($k, \ell = 1, 2, \dots, L$), o vetor de parâmetros dos coeficientes do sistema de equações das médias, β , pode ser estimado por

$$\hat{\beta} = [X_T'(\hat{\rho}^{-1} \otimes I_n)X_T]^{-1} X_T'(\hat{\rho}^{-1} \otimes I_n)^{-1} y_T \quad (4.14)$$

Como o cálculo dos elementos de $\hat{\rho}$ depende do ajuste das equações de regressão de y_1, y_2, \dots, y_L , pode-se usar o estimador dos parâmetros do sistema, $\hat{\beta}$, para obter novos resíduos e calcular novamente $\hat{\rho}$. Assim como nos casos anteriores, recomenda-se repetir o processo iterativo pelo menos mais uma vez.

Cabe lembrar que o método multivariado SUR só produz estimativas mais precisas do que os procedimentos univariados quando as equações têm formas

funcionais diferentes e as correlações, $\rho_{k\ell}$, forem fortes ou moderadas (acima de 0,30, segundo as simulações de Metha e Swamy, 1976).

Nas seções seguintes, o método de estimação aqui descrito, que agrega MQG para correção da heterocedasticidade e o processo multivariado SUR de Zellner, será denotado por MQG-SUR.

4.3 - O SISTEMA DE EQUAÇÕES DAS VARIÂNCIAS

No capítulo 3, discutiram-se as vantagens de se estabelecer as equações de regressão das variâncias da seguinte forma:

$$\log(v_{\ell i} + c\bar{v}_{\ell}) = z_{\ell i}'\theta_{\ell} + u_{\ell i}, \quad \ell = 1, 2, \dots, L; i = 1, 2, \dots, M \quad (4.15)$$

onde: $v_{\ell i}$ é uma estimativa da variância, baseada na variância amostral e no resíduo quadrático médio, referente ao i -ésimo ponto experimental e à ℓ -ésima equação (ver seção 3.4);

\bar{v}_{ℓ} é a média aritmética dos $v_{\ell i}$ da ℓ -ésima equação;

c é uma constante positiva, como definido na seção 3.3;

$z_{\ell i}$ é um vetor formado pela i -ésima linha da matriz de planejamento Z_{ℓ} , referente à ℓ -ésima equação do sistema de regressão das variâncias;

θ_{ℓ} é o vetor de parâmetros da ℓ -ésima equação do sistema das variâncias e

$u_{\ell i}$ é o erro aleatório relativo ao i -ésimo ponto experimental e à ℓ -ésima equação.

Quando o número de replicações é igual ou superior a cinco, pode-se simplificar o sistema de equações (4.16), tomando-se $c = 0$ e variâncias amostrais sobre as replicações, $S_{\ell i}^2$, no lugar dos $v_{\ell i}$, conforme discutido no capítulo anterior. Por outro lado, em experimentos não replicados, usam-se os resíduos quadráticos no lugar dos $v_{\ell i}$.

Uma facilidade no sistema de equações das variâncias é que, com a transformação logarítmica, as variâncias dos erros de cada equação são aproximadamente constantes, especialmente quando $c = 0$. Com isto, pode-se realizar o processo de estimação por MQO. Neste contexto, a estimação multivariada pode ser feita pelo modelo clássico de Zellner, conforme discutido na seção 2.3.2.

Contudo, na maioria das vezes, a estimação é realizada de forma eficiente, tomando-se cada equação da variância isoladamente. Isto porque a magnitude das correlações entre os termos de erro das variâncias é, em geral, bem menor do que a magnitude das correlações entre os termos de erro das respostas, conforme será demonstrado a seguir.

Admita-se, inicialmente, que as variâncias se relacionam com os fatores do processo através de modelos lineares aditivos (sem transformação logarítmica). Desta forma, a matriz de variâncias e covariâncias dos termos de erro das respostas e dos termos de erro das variâncias coincidem com o segundo e o quarto momento das respostas, respectivamente. Sob a hipótese de distribuição normal multivariada, a relação entre o segundo e o quarto momento é conhecida (ver, por exemplo, Press, 1972, p. 107). Para simplificar a exposição, os resultados serão enunciados para o caso bivariado.

Seja $(y_1, y_2)'$ um vetor aleatório normal bivariado com valor esperado $(\mu_1, \mu_2)'$ e matriz de variâncias e covariâncias $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$, então os termos de erro das variâncias têm as seguintes variâncias e covariância:

$$a) \text{Var}\{u_k\} = \text{Var}\{(y_k - \mu_k)^2\} = 2\sigma_{kk}^2 \quad (k = 1, 2) \text{ e}$$

$$b) \text{Cov}\{u_1, u_2\} = \text{Cov}\{(y_1 - \mu_1)^2, (y_2 - \mu_2)^2\} = 2\sigma_{12}^2.$$

Em decorrência de (a) e (b), a correlação entre os termos de erro das variâncias é dada por

$$\rho(u_1, u_2) = \text{Corr}\{(y_1 - \mu_1)^2, (y_2 - \mu_2)^2\} = \frac{2\sigma_{12}^2}{\sqrt{2\sigma_{11}^2}\sqrt{2\sigma_{22}^2}} = \rho_{12}^2 \quad (4.16)$$

onde ρ_{12} é o coeficiente de correlação entre y_1 e y_2 . Como $-1 \leq \rho_{12} \leq 1$, tem-se que $|\rho(u_1, u_2)| \leq |\rho_{12}|$. A figura 4.1 ilustra a relação entre ρ_{12} e $\rho(u_1, u_2)$, onde fica evidente que, em termos absolutos, a correlação entre as variâncias é, em geral, menor do que a correlação entre as respostas.

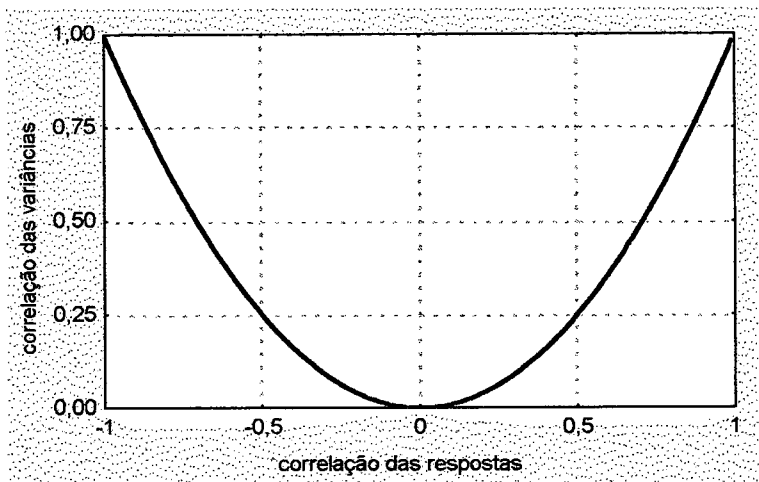


Fig. 4.1 - Relação entre a correlação das variâncias e a correlação das respostas.

Para avaliar o efeito da transformação logarítmica sobre as correlações, levou-se a cabo um estudo de simulação do tipo Monte Carlo. Consideraram-se duas variáveis aleatórias com distribuição normal bivariada de vetor de médias nulo, variâncias unitárias e correlação ρ_{12} . A correlação foi fixada, em cada rodada, num dos seguintes níveis: 0,4, 0,6, 0,8 e -0,8. As equações das variâncias foram estabelecidas através da transformação logarítmica padrão sobre as variâncias amostrais, isto é, $\log(S_{xi}^2)$. O número de replicações, m , usado no cálculo de cada variância amostral, foi fixado em três níveis: 2, 4 e 6. Em cada situação, geraram-se 10.000 amostras para cada resposta. Pares de observações univariadas foram transformados em vetores bivariados utilizando-se a operação inversa da técnica de componentes principais (ver, por exemplo, Johnson e Wichern, 1992, p. 356). A tabela 3.1 mostra os resultados.

Tabela 4.1 - Valores de correlação entre o logaritmo de duas variâncias amostrais.

correlação entre as respostas ρ_{12}	correlação entre as var. amostrais $\rho(u_1, u_2)$	correlação entre os log das var. amostrais		
		$m = 2$	$m = 4$	$m = 6$
0,4	0,16	0,07	0,11	0,14
0,6	0,36	0,18	0,30	0,31
0,8	0,64	0,36	0,53	0,58
-0,8	0,64	0,38	0,53	0,59

Os valores da segunda coluna da tabela 4.1 foram obtidos teoricamente pela relação (4.16), mas os resultados simulados não diferiram desses valores em mais do que um centésimo. As três últimas colunas referem-se aos resultados da simulação propriamente dita. Nota-se que, em todas as situações, a correlação diminuiu após a

transformação logarítmica, especialmente quando as variâncias amostrais foram calculadas com apenas duas observações ($m = 2$).

Com os resultados da simulação, parece razoável admitir que só é recomendável usar o método SUR de Zellner nas equações das variâncias quando o número de replicações for superior a dois e as correlações entre os termos de erro de y_1, y_2, \dots, y_L forem, em valores absolutos, bastante altas. Caso contrário, recomenda-se efetuar a estimação dos parâmetros do sistema das variâncias considerando cada equação isoladamente.

4.4 - ALGORITMO PARA A ESTIMAÇÃO CONJUNTA DOS PARÂMETROS

Esta seção apresenta as linhas gerais para a estimação dos parâmetros dos sistemas de equações das médias e das variâncias. Ressalta-se que cada problema tem as suas peculiaridades, que podem exigir adequações dos passos descritos a seguir. Admite-se que já se tenha identificado os principais efeitos associados a cada resposta.

PASSO 1 - Estimam-se, preliminarmente, os parâmetros do sistema de equações das respostas (ou das médias) pelo processo usual dos mínimos quadrados (MQO), aplicado a cada equação isoladamente. Analisam-se os resíduos para verificar a adequação do ajuste e, se necessário, efetuam-se transformações nas variáveis. Fazem-se, para cada equação, testes estatísticos de validade do modelo e de falta de ajuste (testes F).

PASSO 2 - Calculam-se as variâncias amostrais ($S_{\ell i}^2$, $\ell = 1, 2, \dots, L$ e $i = 1, 2, \dots, M$). O cálculo é feito para cada resposta e em todos os pontos experimentais ensaiados com replicações. Quando o experimento tem menos do que cinco replicações ($m < 5$), calculam-se também os resíduos quadráticos médios ($\hat{e}_{\ell i}^2$, $\ell = 1, 2, \dots, L$ e $i = 1, 2, \dots, M$) e adotam-se as medidas agregadas de variâncias: $v_{\ell i} = \gamma \hat{e}_{\ell i}^2 + (1-\gamma)S_{\ell i}^2$, conforme apresentação no capítulo anterior. Em experimentos não replicados, usam-se apenas os resíduos quadráticos.

PASSO 3 - Fazem-se as estimações dos parâmetros do sistema de equações das variâncias, utilizando-se transformação logarítmica e MQO em cada equação. O

estabelecimento das equações de regressão pode ser feito com as técnicas discutidas no capítulo 3. Realizam-se testes F para verificar a significância de cada equação.

PASSO 4 - Ainda considerando cada equação isoladamente, estimam-se os coeficientes das equações das médias através da técnica dos mínimos quadrados generalizados (MQG), usando-se como peso das observações o inverso das variâncias preditas nas equações estimadas no passo 3. Nas equações onde o teste F do passo 3 não acusou significância, mantém-se a estimação por MQO. Quando se usa MQG, reavaliam-se as medidas de variância $v_{\ell i}$ ($\ell = 1, 2, \dots, L$ e $i = 1, 2, \dots, M$).

PASSO 5 - Calculam-se os resíduos das regressões estimadas no passo 4. Estes resíduos, quando utilizado o método MQG, devem ser calculados com as variáveis transformadas, isto é, divididas pelos correspondentes desvios padrão. Com estes resíduos, calculam-se as correlações ($\hat{\rho}_{k\ell}$, $k \neq \ell = 1, 2, \dots, L$).

Os passos seguintes tratam do processo de estimação multivariada propriamente dito e só devem ser efetuados se:

- a) as correlações $\hat{\rho}_{k\ell}$ forem, em geral, significantes e superiores a 0,30 e
- b) os conjuntos de variáveis independentes das equações não forem todos iguais (ou muito parecidos).

PASSO 6 - Monta-se o vetor de observações ampliado, y , com as nL observações referentes a todas as respostas. Monta-se, também, a matriz de planejamento X , de dimensão $nL \times P$, formada pelas matrizes de planejamento de cada equação e estabelece-se o sistema SUR (ver expressão 4.3).

PASSO 7 - Estimam-se os parâmetros das equações de regressão das médias, aplicando-se MQG sobre o sistema SUR. Neste processo, faz-se uso dos pesos, conforme descrito no passo 4, como também, da matriz de correlações construída no passo 5 (ver expressão 4.14).

Com os resíduos, que podem ser calculados em relação ao modelo do passo 7, pode-se obter novas equações para as variâncias (passo 3) e estimar novamente a matriz de correlações (passo 5) e, assim, repetir o processo iterativamente. Recomenda-se repetir os passos 3, 5, 6 e 7 mais uma vez.

O algoritmo MQG-SUR também pode ser apresentado sem o passo 4, mas a forma mais detalhada, como descrito, permite comparar os três métodos de estimação (MQO para cada equação, MQG para cada equação e MQG-SUR para o sistema). Isto possibilita verificar até que ponto a correção da heterocedasticidade altera os resultados do MQO e, também, quão diferente tornam-se as estimativas quando se considera o sistema como um todo.

Outra alteração possível é estimar o sistema de equações de variância também de forma conjunta. Neste caso, pode-se incluir mais três passos: 8, 9 e 10, análogos aos passos 5, 6 e 7, mas substituindo as equações de regressão das médias por equações de regressão das variâncias. Conforme discutido na seção anterior, esta ampliação do algoritmo somente deverá levar a estimativas mais precisas nos casos em que as correlações entre os termos de erro das respostas forem altas e houver mais de duas replicações.

Observa-se que os estimadores sugeridos no presente algoritmo são consistentes. Desta forma, as propriedades assintóticas dos testes estatísticos sob o modelo SUR permanecem válidas, permitindo suas aplicações em problemas com amostras grandes.³

4.5 - ESTUDOS DE SIMULAÇÃO

Para avaliar o procedimento proposto, realizaram-se algumas simulações do tipo Monte Carlo em situações típicas, usando-se sempre duas respostas correlacionadas segundo um coeficiente ρ_{12} . Adotaram-se dois tipos de experimentos, descritos a seguir.

Experimento 2⁴. Realizado com 4 fatores, sendo cada um deles ensaiados a dois níveis e duas replicações em cada cruzamento de níveis. Os fatores, representados por x_1, x_2, x_3 e x_4 , foram codificados com os valores ± 1 . As equações das médias foram geradas pelas equações

$$\mu_{y_i} = \beta_0^{(i)} + \beta_1^{(i)}x_1 + \beta_2^{(i)}x_2 + \beta_3^{(i)}x_3 + \beta_4^{(i)}x_4 + \beta_{12}^{(i)}x_1x_2 + \beta_{23}^{(i)}x_2x_3 + \beta_{34}^{(i)}x_3x_4 \quad (4.17)$$

³ De maneira informal, diz-se que um estimador é *consistente* quando ele tende para o verdadeiro parâmetro na medida em que o tamanho da amostra (n) tende para infinito. Propriedades *assintóticas* são obtidas fazendo-se $n \rightarrow \infty$ nas expressões matemáticas.

e

$$\mu_{y_2} = \beta_0^{(2)} + \beta_1^{(2)}x_1 + \beta_2^{(2)}x_2 + \beta_3^{(2)}x_3 \quad (4.18)$$

com $\beta_0^{(\ell)} = 0$, $\beta_j^{(\ell)} = 1$ e $\beta_{ij}^{(\ell)} = 0,5$ ($\ell = 1, 2$ e $i < j = 1, 2, 3, 4$). Os super-índices dos coeficientes, que indicam o número da equação, serão suprimidos quando não houver risco de confusão. Ambas as equações das variâncias foram geradas pelo mesmo padrão, dada pela expressão

$$\sigma_{y_1}^2 = \sigma_{y_2}^2 = \exp\{\theta_0 + \theta_1x_1 + \theta_2x_2 + \theta_3x_3\} \quad (4.19)$$

com $\theta_0 = 0$, $\theta_1 = 0,693$, $\theta_2 = 0,549$ e $\theta_3 = 0,347$.

Os termos aleatórios foram gerados pela distribuição normal bivariada de variâncias unitárias e correlação ρ_{12} . Utilizaram-se duas situações: $\rho_{12} = 0,4$ e $\rho_{12} = 0,8$. As equações das médias, as equações das variâncias e os termos aleatórios foram agregados conforme as expressões gerais (4.1) e (4.2).

Experimento CCD. Realizado com três fatores, ensaiados com 8 pontos fatoriais nas combinações de $x_i = \pm 1$ ($i = 1, 2, 3$), 8 pontos axiais em $x_i = \pm 1,682$ e $x_j = 0$ ($j \neq i = 1, 2, 3$) e 4 pontos centrais $x_i = 0$ ($i = 1, 2, 3$). Esta estrutura foi repetida três vezes (3 replicações), totalizando 60 ensaios. As equações das médias foram geradas pelas equações

$$\mu_{y_1} = \beta_0^{(1)} + \beta_1^{(1)}x_1 + \beta_2^{(1)}x_2 + \beta_3^{(1)}x_3 + \beta_{11}^{(1)}x_1^2 + \beta_{22}^{(1)}x_2^2 + \beta_{33}^{(1)}x_3^2 + \beta_{12}^{(1)}x_1x_2 + \beta_{13}^{(1)}x_1x_3 + \beta_{23}^{(1)}x_2x_3 \quad (4.20)$$

$$\mu_{y_2} = \beta_0^{(2)} + \beta_1^{(2)}x_1 + \beta_2^{(2)}x_2 + \beta_3^{(2)}x_3 \quad (4.21)$$

com $\beta_0^{(\ell)} = 0$, $\beta_j^{(\ell)} = 1$ e $\beta_{ij}^{(\ell)} = 0,5$ ($\ell = 1, 2$ e $i \leq j = 1, 2, 3$).

As equações das variâncias e os termos aleatórios foram gerados da mesma forma do que no experimento anterior, assim como a especificação da correlação e a agregação das equações e dos termos aleatórios para a composição das respostas bivariadas.

Conforme a teoria de sistemas SUR (ver, por exemplo, Srivastava e Giles, 1987), os valores dos coeficientes das equações das médias não devem interferir no desempenho do método. Já os coeficientes das equações das variâncias devem influenciar o desempenho do método MQG, conforme discutido no capítulo anterior. Por outro lado, a correlação entre os termos de erro das duas equações e o número de

diferentes variáveis independentes nas duas equações são fatores determinantes no sucesso do método multivariado SUR.

No processo de estimação, os modelos das variâncias foram especificados com a transformação logarítmica modificada, adotando-se $c = 0,06$ no experimento 2⁴ e $c = 0,08$ no CCD.⁴ A estimação dos parâmetros foi realizada pelos três seguintes procedimentos:

- 1) o método SUR assumindo homocedasticidade, conforme descrito no capítulo 2;
- 2) o método MQG univariado e
- 3) o algoritmo descrito na seção 4.4 (MQG-SUR).

Para cada um dos procedimentos, avaliou-se a eficiência em relação ao método MQO univariado. Os resultados baseiam-se em 1.000 realizações simuladas de cada experimento e estão descritos nas figuras 4.2 a 4.9. O programa computacional encontra-se no apêndice.

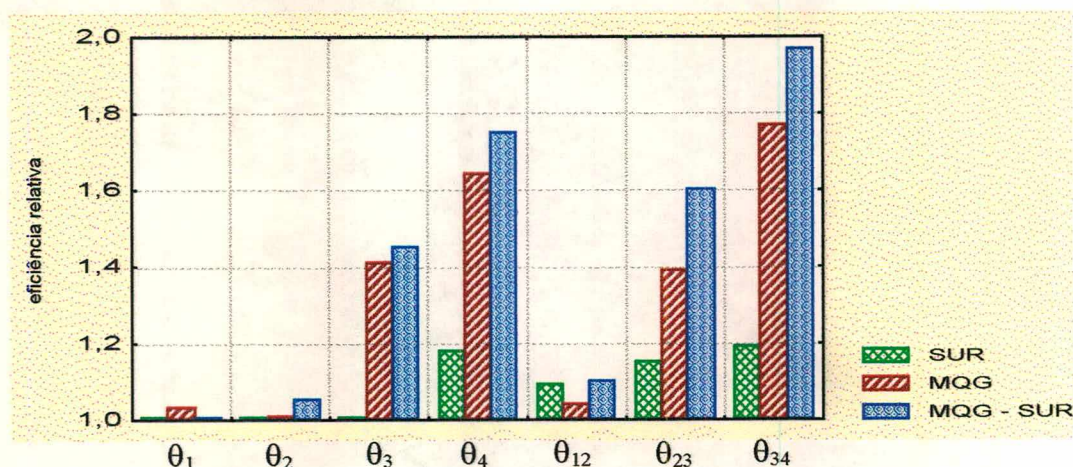


Fig. 4.2 - Eficiência dos métodos SUR, MQG e MQG-SUR em relação ao MQO. Experimento 2⁴, $\rho_{12} = 0,40$, primeira equação.

A figura 4.2 mostra os níveis de eficiência na estimação de cada coeficiente da primeira equação do sistema das médias, expressão (4.17), considerando os dados gerados com $\rho_{12} = 0,40$. Nota-se que o método MQG univariado foi superior ao MQO univariado (eficiência relativa superior à unidade). Isto era esperado porque as equações foram geradas com variâncias desiguais. Nota-se, também, que a eficiência da estimação por MQG em relação ao MQO não foi igual para todos os coeficientes,

⁴ Os valores de c foram calculados pelas expressões (3.21) e (3.28), respectivamente.

como ocorre em vários exemplos descritos na literatura (ver, por exemplo, Chatterjee e Price, 1977, p. 114). O processo de estimação multivariado SUR apresentou pouca melhora em relação ao MQO univariado, quando foi assumido homocedasticidade. Por outro lado, quando se incorporou o tratamento da heterocedasticidade no SUR (MQG-SUR), obteve-se os melhores resultados, com destaque para as estimativas dos coeficientes β_4 , β_{12} , β_{23} e β_{34} , cujos fatores correspondentes não estão na segunda equação. A melhora do SUR - MQG em relação ao MQG univariado, contudo, foi modesta.

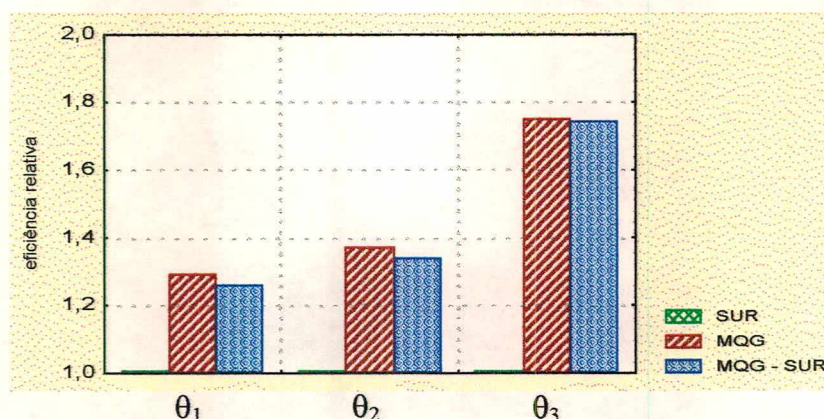


Fig. 4.3 - Eficiência dos métodos SUR, MQG e MQG-SUR em relação ao MQO. Experimento 2⁴, $\rho_{12} = 0,40$, segunda equação.

A figura 4.3 mostra os resultados relativos à segunda equação (expressão 4.18). Observa-se que o método SUR clássico (equações homocedásticas) apresentou a mesma eficiência que o MQO univariado (eficiência relativa igual a unidade), enquanto o MQG univariado e o MQG-SUR apresentaram desempenhos muito parecidos. Este resultado já era esperado até certo ponto, pois no método SUR clássico, quando as variáveis independentes de uma equação B formam um subconjunto das variáveis independentes da outra equação A, o método SUR apresenta, para a equação B, os mesmos resultados do que se obteria estimando-a isoladamente (ver Srivastava e Giles, 1987, p. 104). Os resultados da simulação, neste caso, indicam que esta propriedade do modelo SUR clássico é mantida no algoritmo proposto, pelo menos aproximadamente e nas mesmas condições em que foram feitos os estudos de simulação.

As figuras 4.4 e 4.5 referem-se aos resultados de eficiência relativa das duas equações do experimento 2⁴, com alta correlação entre os termos de erro ($\rho_{12} = 0,80$).

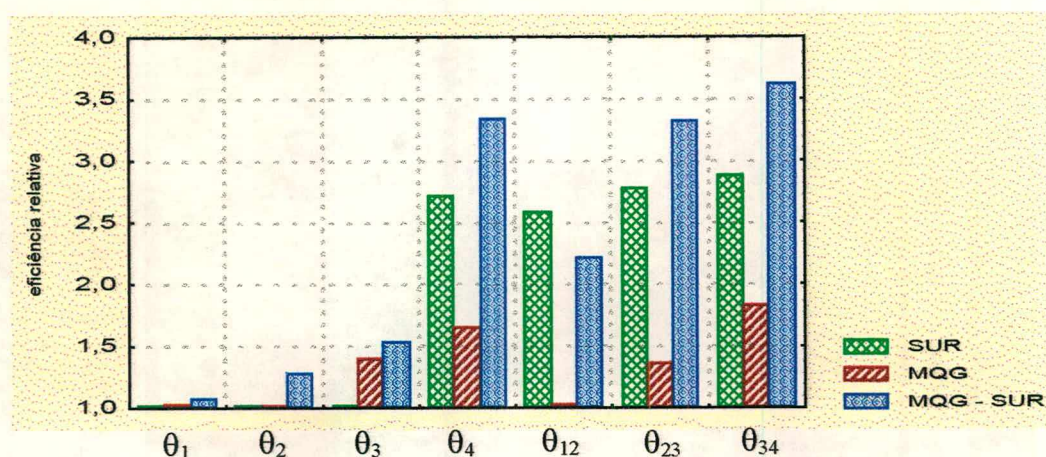


Fig. 4.4 - Eficiência dos métodos SUR, MQG e MQG-SUR em relação ao MQO. Experimento 2⁴, $\rho_{12} = 0,80$, primeira equação.

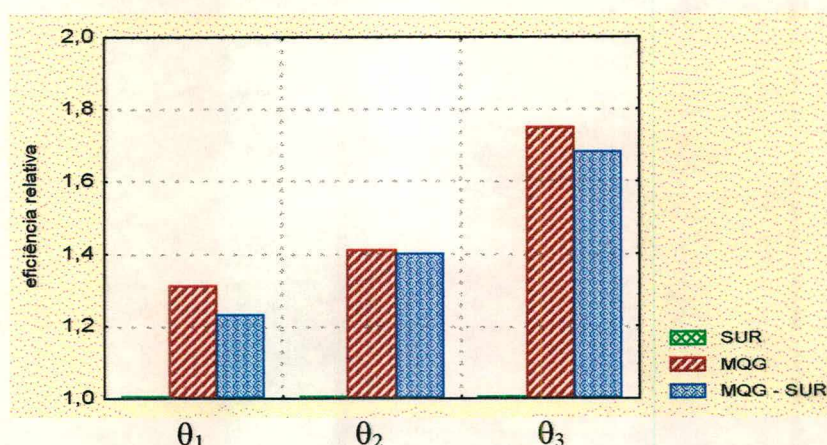


Fig. 4.5 - Eficiência dos métodos SUR, MQG e MQG-SUR em relação ao MQO. Experimento 2⁴, $\rho_{12} = 0,80$, segunda equação.

A figura 4.4, referente à primeira equação do sistema das médias, é apresentada com escala ampliada para contemplar todos os valores de eficiência relativa. O método MQG univariado apresentou praticamente o mesmo desempenho observado na figura 4.2, onde os dados foram gerados com correlação mais fraca. Já o método SUR e, principalmente, o MQG-SUR levaram a grandes ganhos em eficiência nas estimativas dos coeficientes β_4 , β_{12} , β_{23} e β_{34} , cujos fatores correspondentes não estão na segunda equação. É uma situação típica onde o algoritmo proposto deve ser aplicado. Quanto à segunda equação (figura 4.5), cujo lado direito é formado por um subconjunto das variáveis independentes da primeira equação, o processo multivariado teve desempenho levemente inferior ao univariado.

As figuras 4.6 a 4.9 referem-se ao experimento CCD, onde tem-se maior número de pontos experimentais por fator e maior número de replicações. Na teoria

clássica de SUR, o aumento do tamanho da amostra tende a favorecer o método, pois, em termos assintóticos, ele é melhor do que o MQO univariado para qualquer $\rho_{12} \neq 0$.

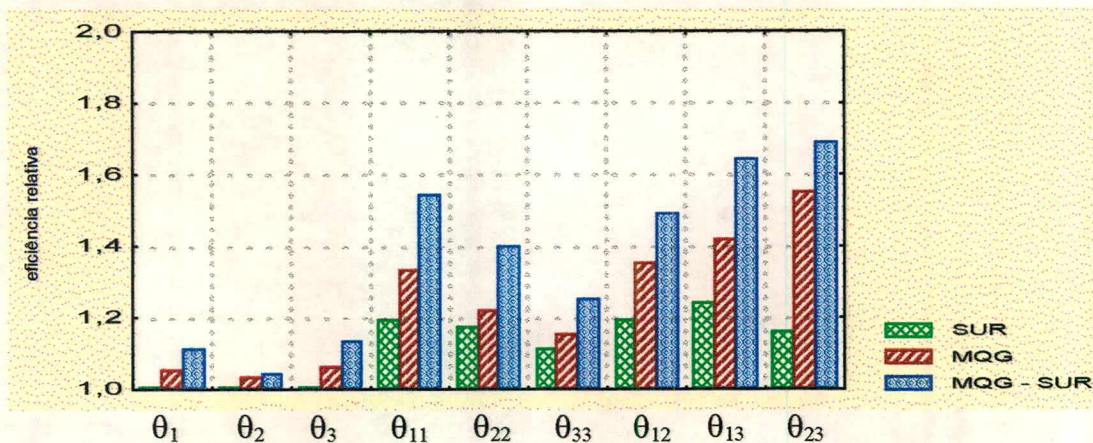


Fig. 4.6 - Eficiência dos métodos SUR, MQG e MQG-SUR em relação ao MQO. Experimento CCD, $\rho_{12} = 0,40$, primeira equação.

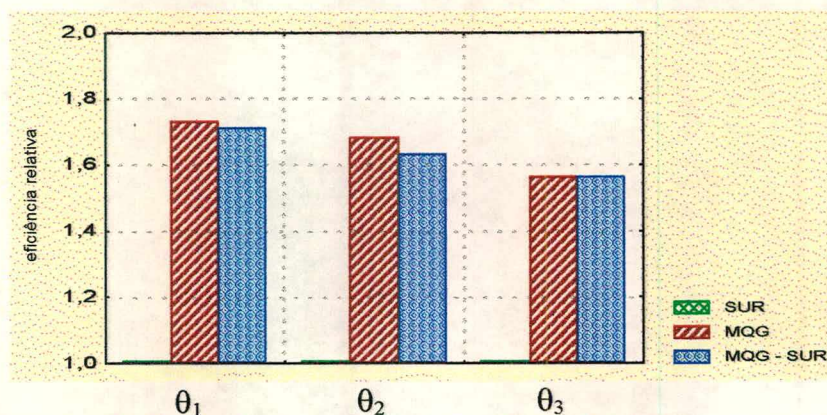


Fig. 4.7 - Eficiência dos métodos SUR, MQG e MQG-SUR em relação ao MQO. Experimento CCD, $\rho_{12} = 0,40$, segunda equação.

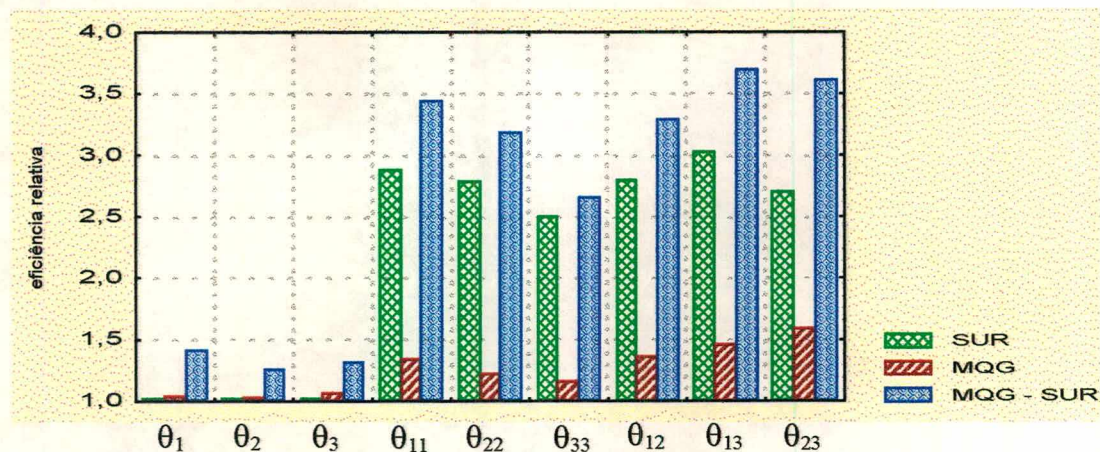


Fig. 4.8 - Eficiência dos métodos SUR, MQG e MQG-SUR em relação ao MQO. Experimento CCD, $\rho_{12} = 0,80$, primeira equação.

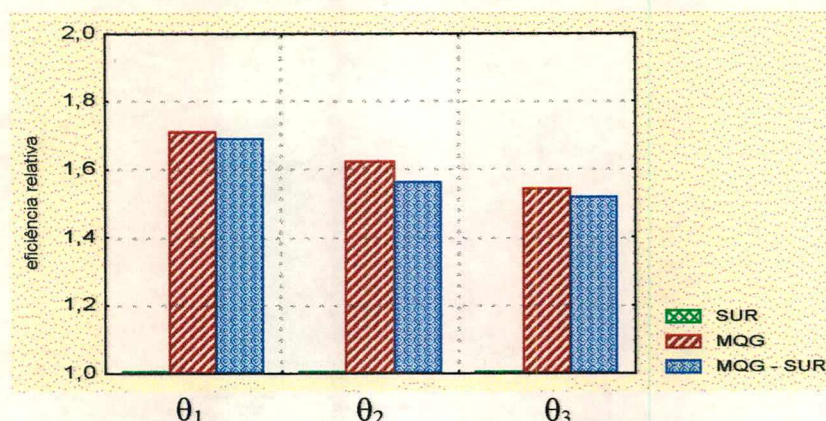


Fig. 4.9 - Eficiência dos métodos SUR, MQG e MQG-SUR em relação ao MQO. Experimento CCD, $\rho_{12} = 0,80$, segunda equação.

De modo geral, os resultados do experimento CCD (figuras 4.6 a 4.9) apresentaram resultados muito parecidos com o experimento 2^4 , notando-se na figura 4.6, uma vantagem levemente mais nítida dos procedimentos SUR e MQG-SUR, quando comparada à figura 4.2.

4.6 - ROBUSTEZ EM RELAÇÃO À HIPÓTESE DE CORRELAÇÕES CONSTANTES

O algoritmo descrito na seção 4.4 utilizou suposições pouco comuns em sistemas de equações aparentemente não relacionadas (SUR), como variâncias desiguais (heterocedasticidade) juntamente com correlações constantes. Em outras palavras, assume-se que os fatores do processo (x_1, x_2, \dots, x_k) podem alterar as variâncias de cada resposta, mas não influenciam as correlações entre as respostas. Como a suposição de correlações constantes é difícil de ser verificada na prática, fez-se um estudo sobre a robustez do algoritmo em relação a possíveis violações desta suposição.

O estudo baseou-se num experimento fatorial 2^4 com 2 replicações e equações das médias geradas pelas equações

$$\mu_{y_1} = \beta_0^{(1)} + \beta_1^{(1)}x_1 + \beta_2^{(1)}x_2 + \beta_3^{(1)}x_3 \quad (4.22)$$

$$\mu_{y_2} = \beta_0^{(2)} + \beta_1^{(2)}x_1 + \beta_4^{(2)}x_4 \quad (4.23)$$

com $\beta_0^{(\ell)} = 0$ e $\beta_j^{(\ell)} = 1$ ($\ell = 1, 2$ e $j = 1, 2, 3, 4$). As equações das variâncias foram especificadas como em (4.19) e os termos de erro foram gerados pela distribuição

normal bivariada de variâncias unitárias e correlação ρ_{12} , definida segundo duas situações:

- correlação constante dada por $\rho_{12} = 0,60$ e
- correlação variando segundo o fator x_1 , com valor $\rho_{12} = 0,40$ quando $x_1 = -1$ e $\rho_{12} = 0,80$ quando $x_1 = +1$.

Observa-se que a situação (b) viola a suposição de que a correlação não se altera com os fatores do processo. Foram realizadas 1000 simulações do experimento em cada situação. A figura 4.10 mostra a eficiência do algoritmo proposto (MQG-SUR), em relação ao MQO univariado, sob os dados gerados segundo a situação (a) (correlação constante) e sob os dados gerados conforme a situação (b) (correlações diferentes). No segundo caso, incluiu-se, também, a eficiência do MQG univariado em relação ao MQO univariado.

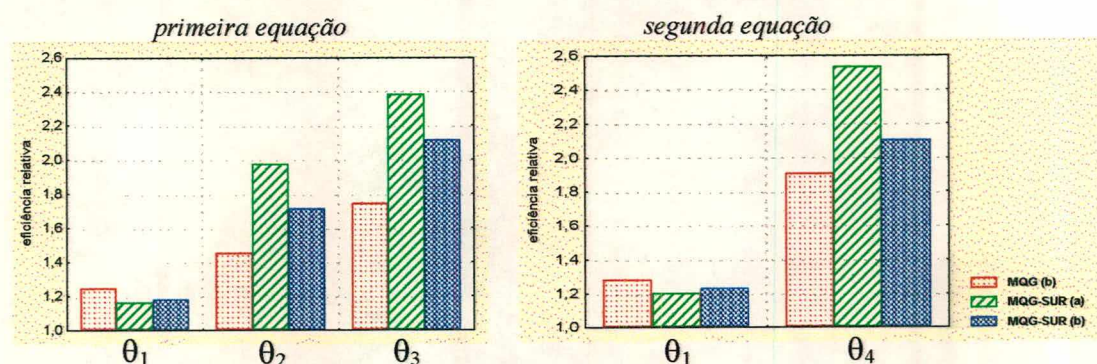


Fig. 4.10 - Eficiência dos métodos MQG e MQG-SUR em relação ao MQO. Dados gerados de duas formas: (a) ρ_{12} fixo em 0,60 e (b) ρ_{12} variando de 0,40 a 0,80.

Observa-se, na figura 4.10, que mesmo violando fortemente a suposição de correlações constantes, o algoritmo MQG-SUR ainda funciona razoavelmente bem, apresentando desempenho superior ao método MQG univariado. Em outras palavras, o método MQG-SUR, proposto na seção 4.4, pode ser considerado robusto (ou pouco sensível) a violação da suposição de correlações constantes.

De modo geral, as simulações realizadas nestas duas últimas seções credenciam o uso do algoritmo de estimação proposto na seção 4.4, designado por MQG-SUR, para as situações em que as respostas:

- tenham variâncias que se alteram significativamente na região experimental;
- tenham entre si correlações moderadas ou fortes e
- estão associadas aos fatores do processo por diferentes formas funcionais.

ESTRATÉGIAS DE ANÁLISE E ESTUDOS DE CASO

O presente capítulo enfatiza as situações de estudos experimentais onde se tem interesse em construir modelos tanto para a média quanto para a variância da variável em estudo, incluindo o caso de múltiplas respostas (multivariado). Os procedimentos são ilustrados com três estudos de caso extraídos da literatura especializada.

5.1 - ESTRATÉGIAS DE ANÁLISE

Na construção de modelos para a variância, procura-se juntar os procedimentos discutidos no capítulo 3, descrevendo-se uma estratégia mais geral. A estratégia baseia-se na busca de uma estrutura de variância baseada nas variações não explicadas pelas médias de cada ponto experimental (método S2) e/ou pelo modelo da resposta (método RQ). Esta estrutura geralmente é o próprio modelo da variância. Somente nos casos em que componentes de ruído são incluídas no modelo da resposta, a estrutura de variância, baseada nos termos de erro, passa a ser apenas um termo adicional do modelo da variância (ver seção 2.2.6).

A discussão inicia-se com experimentos não replicados, enfatizando-se os planejamentos fatoriais (tipo 2^k , 3^k ou mistos, completos ou fracionados). Nestes experimentos, os pontos experimentais são ensaiados de tal forma que a projeção sobre cada fator forma um sistema ortogonal, permitindo a agregação de fatores não significativos. Esta agregação forma *pseudo-replicações*, as quais permitem avaliar melhor os efeitos de dispersão. Posteriormente, discute-se sobre a construção de modelos em experimentos replicados, dando-se ênfase à transformação logarítmica modificada e à agregação de variâncias amostrais com resíduos quadráticos. Por fim, consideram-se as situações de múltiplas respostas.

5.1.1 - Experimentos não replicados

Consideram-se, inicialmente, experimentos fatoriais não replicados, onde normalmente existem fatores que podem ser negligenciados. O passo inicial é identificar possíveis fatores que alteram a média do processo (efeitos de locação) e fatores que alteram a variância do processo (efeitos de dispersão), onde sugerem-se os seguinte passos:

- a) com os dados experimentais, estimam-se os possíveis efeitos, apresentando-os num gráfico de probabilidade normal (gráfico normal dos efeitos de locação);
- b) com os efeitos identificados no item (a), ajusta-se um modelo preliminar por MQO (modelo inicial da média do processo);
- c) faz-se uma análise exploratória dos resíduos para verificar se a métrica da resposta é adequada e, no caso de se precisar realizar alguma transformação nos dados, estes três primeiros passos devem ser repetidos;
- d) calculam-se os resíduos quadráticos ($\hat{\epsilon}_i^2$, $i = 1, 2, \dots, n$) e efetua-se a transformação $\log(\hat{\epsilon}_i^2 + \bar{\epsilon}^2)$, onde $\bar{\epsilon}^2$ é a média dos resíduos quadráticos e
- e) usam-se as transformadas dos resíduos como se fossem observações da variância e estimam-se os possíveis efeitos, apresentando-os num gráfico de probabilidade normal (gráfico normal dos efeitos de dispersão).¹

Os gráficos normais dos efeitos construídos nos itens (a) e (e) permitem identificar os principais efeitos que alteram a média e a variância do processo, possibilitando uma análise preliminar sobre o estudo experimental em questão. Em geral, nos experimentos desta natureza, não se constróem modelos, especialmente para a variância do processo, devido à falta de informações confiáveis para este fim.

¹ No capítulo 3, discutiu-se sobre a transformação logarítmica modificada, onde se considerava a adição de uma proporção da média dos resíduos quadráticos, $c\bar{\epsilon}^2$, no argumento da função logarítmica. No presente contexto, optou-se por tomar $c = 1$, considerando-se que na identificação dos efeitos a análise é feita de forma relativa e, como visto na seção 3.3.5, o vício devido à transformação é aproximadamente constante para os vários efeitos ortogonais. Desta forma, a utilização de um valor alto para c deve discernir melhor os efeitos significativos no gráfico normal dos efeitos.

No capítulo 2, também se argumentou sobre a correção dos graus de liberdade perdidos na estimação dos parâmetros do modelo da média. Esta correção pode ser efetuada nos resíduos (ver seção 2.2.4), mas ela é mais relevante no processo de estimação do que na fase de identificação.

Contudo, se a construção dos modelos for fundamental e não houver a possibilidade de ampliar o experimento, descreve-se, a seguir, uma possível estratégia.

Quando o experimento é do tipo fatorial, a construção do modelo da variância é facilitada com a agregação de fatores que não têm efeitos significativos de dispersão, formando *pseudo-replicações*. Nos pontos experimentais formados pelos níveis dos fatores de dispersão significativos, calculam-se os resíduos quadráticos médios e os valores da estatística $\log(\hat{e}_i^2 + c\bar{e}^2)$, onde \hat{e}_i^2 é o resíduo quadrático médio do i-ésimo ponto experimental, \bar{e}^2 é a média de todos os resíduos quadráticos e c é avaliado segundo uma das expressões apresentadas a seguir.²

Para experimentos fatoriais do tipo 2^k :

$$c = \exp\{-4,22 - 1,70A - 0,58D - 1,85F + 0,34A^2 + 0,39D^2 + 0,54F^2 + 0,59AD + 0,20AF + 0,45DF\} \quad (5.1)$$

onde: $A = \frac{k-5}{2}$, sendo k o número de fatores (já descontados eventuais fracionamentos);

$$D = \frac{m-3}{2}, \text{ sendo } m \text{ o número de replicações (ou } pseudo\text{-replicações) e}$$

$$F = \frac{\theta_* - 0,7}{0,4}, \text{ sendo } \theta_* \text{ o maior coeficiente do modelo da variância}$$

Para experimentos fatoriais do tipo 3^k :

$$c = \exp\{-2,52 - 1,92A - 0,86D - 2,04F + 0,12A^2 + 0,16D^2 + 0,46F^2 + 0,52AD + 0,28AF + 0,19DF\} \quad (5.2)$$

onde: $A = k - 3$, sendo k o número de fatores (já descontados eventuais fracionamentos);

$$D = m - 2, \text{ sendo } m \text{ o número de replicações (ou } pseudo\text{-replicações) e}$$

$$F = \frac{\theta_* - 0,7}{0,4}, \text{ sendo } \theta_* \text{ o coeficiente de maior magnitude do modelo da variância.}$$

Para experimentos do tipo CCD:

$$c = \exp\{-1,94 - 0,76A - 0,98D - 2,21F - 0,05A^2 + 0,45D^2 + 0,58F^2 + 0,29AD - 0,14AF + 0,22DF\} \quad (5.3)$$

onde A , D e F são definidos como em (5.2).

Observa-se que as expressões (5.1), (5.2) e (5.3) são funções do número de fatores (k), número de replicações ou *pseudo-replicações* (m) e do coeficiente de maior magnitude do modelo da variância (θ_*). Este último não é conhecido, mas

² Nesta etapa, é recomendável considerar os resíduos corrigidos, conforme descrito na seção 2.2.4. As expressões (5.1), (5.2) e (5.3) foram desenvolvidas no capítulo 3.

pode-se usar alguma estimativa feita previamente. Outra opção é usar $\theta_* = 0,7$, que corresponde a uma razão entre os desvios padrão de 2 (dois). Quando se obtém as estimativas dos coeficientes, pode-se reavaliar o valor de c e repetir o processo.

Para se estabelecer o modelo da variância, os valores da estatística $\log(\hat{e}_i^2 + c\bar{e}^2)$ são usados como se fossem observações num modelo de regressão. Estimativas iniciais podem ser obtidas com a aplicação de MQO, resultando num modelo do tipo

$$\log(\hat{\sigma}_y^2) = \hat{\theta}_0 + \hat{\theta}_1 x_1 + \dots \quad (5.4)$$

ou

$$\hat{\sigma}_y^2 = \exp\{\hat{\theta}_0 + \hat{\theta}_1 x_1 + \dots\} \quad (5.5)$$

Devido à transformação logarítmica, o fator constante da equação (5.5), $\hat{\sigma}_0^2 = \exp\{\hat{\theta}_0\}$, é estimado pelo método MQO de forma viciada. Sugere-se, então, substituí-lo pela estimativa do erro quadrático médio (ou resíduo quadrático médio, \bar{e}^2). Esta estimativa deve ser feita com a devida correção da perda dos graus de liberdade do modelo da média.³

Quando o teste F do modelo da variância acusar significância estatística, recomenda-se refazer as estimativas do modelo da média, usando o método dos mínimos quadrados generalizados (MQG). Por este método, cada observação é ponderada pelo inverso de sua variância, sendo que as variâncias podem ser previstas pela expressão (5.5).⁴ Com o novo modelo da média, tem-se novos resíduos, onde pode-se estimar novamente os parâmetros do modelo da variância (processo iterativo).

5.1.2 - Experimentos com replicações

Por simplicidade, a discussão será direcionada para experimentos que tenham replicações em todos os pontos experimentais ensaiados. Porém, os métodos aqui

³ Quando se está trabalhando com os resíduos corrigidos, a estimativa do erro quadrático médio é simplesmente a média dos resíduos quadráticos, caso contrário, deve-se descontar, no denominador da expressão da média, o número de parâmetros estimados no modelo da média do processo.

⁴ Computacionalmente, o método MQG pode ser feito por uma rotina de MQO. Antes de proceder as estimações, divide-se as observações da variável dependente e de todos os fatores (inclusive o vetor de elementos unitários) pelo inverso do desvio padrão da correspondente observação. A aplicação de MQO nestas variáveis transformadas equivale ao MQG nas variáveis originais.

discutidos poderão ser adaptados para experimentos com replicações em apenas alguns pontos.

Havendo replicações, pode-se ter uma estimativa preliminar da variância do processo, admitindo-a inicialmente constante. Esta estimativa permite identificar o modelo da média por testes estatísticos baseados na razão F. Este procedimento é mais formal do que o gráfico normal dos efeitos e, por isto, recomenda-se o seu uso quando possível.

Nos experimentos com cinco ou mais replicações, o modelo da variância pode ser construído a partir do logaritmo da variância amostral. Em cada ponto experimental, calcula-se a variância amostral e efetua-se a transformação logarítmica: $\log(S_i^2)$, $i = 1, 2, \dots, M$, onde M é o número de pontos experimentais. Com esta estatística, pode-se, primeiramente, identificar os efeitos de dispersão significativos através de um gráfico normal dos efeitos e, com o modelo especificado em termos destes efeitos, aplicar MQO para se obter um modelo para a variância. Se o teste F do modelo da variância acusar significância estatística, o modelo da média pode ser estimado por MQG. Neste caso, não se precisa fazer o processo iterativo porque o modelo da variância independe do modelo da média.

Na análise dos dados de experimentos que têm entre duas e quatro replicações, a estatística $\log(S_i^2)$ é ineficiente para avaliar a dispersão no i -ésimo ponto experimental ($i = 1, 2, \dots, M$). Neste caso, recomenda-se o seguinte procedimento:

- a) ajusta-se um modelo preliminar para a média do processo através da aplicação de MQO sobre as observações;
- b) calculam-se o nível descritivo de significância do teste F de falta de ajuste (p_F) e os resíduos (\hat{e}_i , $i = 1, 2, \dots, n$, onde n é o número total de observações);⁵
- c) calculam-se as seguintes medidas de variância (ver detalhes na seção 3.4):

$$v_i = \gamma \hat{e}_i^2 + (1-\gamma)S_i^2 \quad (i = 1, 2, \dots, M),$$

sendo: S_i^2 a variância amostral do i -ésimo ponto experimental;

⁵ Se o número de observações não for bastante superior ao número de parâmetros do modelo da média, recomenda-se efetuar correções nos resíduos (ver seção 2.2.4).

\hat{e}_i^2 o resíduo quadrático médio do i -ésimo ponto experimental e

$$\gamma = \begin{cases} 1 & \text{se } p_F \geq 0,50 \\ (0,50 - p_F)/0,49 & \text{se } 0,01 \leq p_F < 0,50 \\ 0 & \text{se } p_F \leq 0,01 \end{cases}$$

- d) calculam-se $\log(v_i + \bar{v})$, onde \bar{v} é a média dos v_i ($i = 1, 2, \dots, M$);
- e) calculam-se os efeitos em termos da estatística $\log(v_i + \bar{v})$ e apresentam-os num gráfico de probabilidade normal para identificar os efeitos de dispersão significativos;
- f) com o modelo da variância especificado, estimam-se os parâmetros usando $\log(v_i + c\bar{v})$ como se fossem observações da variância num modelo de regressão (o valor de c pode ser avaliado pela expressão 5.1, 5.2 ou 5.3) e
- g) se o teste F do modelo da variância acusar significância estatística, reavaliam-se as estimativas do modelo da média com a aplicação do método MQG, onde os pesos correspondem ao inverso das predições do modelo da variância.

Os passos de (b) a (g) podem ser repetidos mais uma vez para melhorar a qualidade das estimativas. Além disso, o fator constante do modelo exponencial da variância pode ser substituído por \bar{v} (média aritmética dos v_i , $i = 1, 2, \dots, M$), pois, conforme discutido previamente, a estimativa deste fator por MQO ou MQG é viciada devido a transformação logarítmica.

5.1.3 – Construção dos modelos

A figura 5.1 apresenta um fluxograma com as principais fases para construção de modelos para a média e a variância da resposta. Este fluxograma deve servir apenas como uma orientação geral, considerando as peculiaridades de cada aplicação particular.

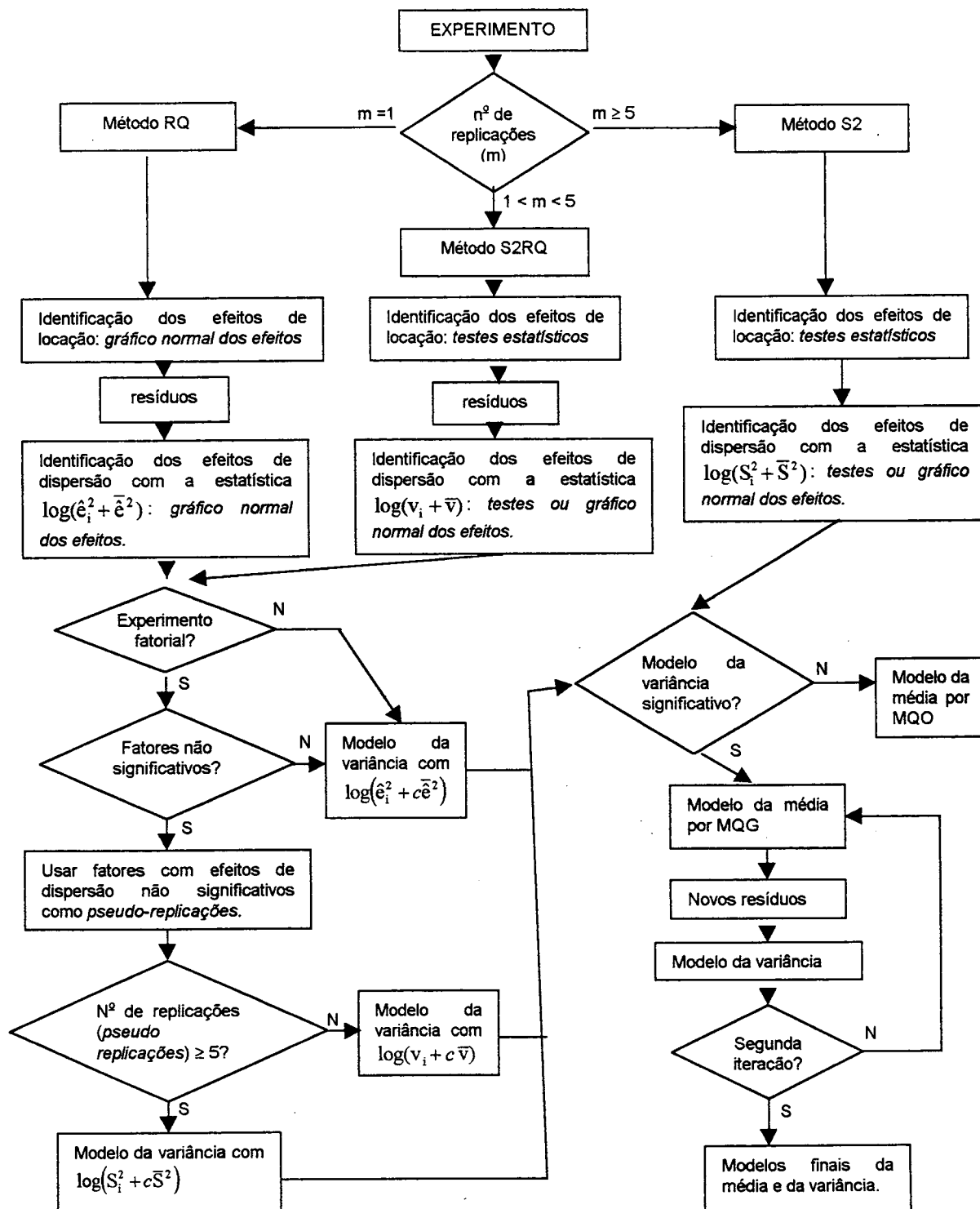


Fig. 5.1 – Estratégia geral para construir modelos para a média e para a variância da resposta.

5.1.4 - Experimentos com múltiplas respostas

Uma situação mais complexa, mas comum na prática, é quando o experimento produz várias respostas. Uma forma de análise é proceder a identificação dos fatores e construir modelos preliminares para cada resposta isoladamente, conforme discutido nas seções anteriores. A seguir, observam-se as seguintes condições:

- a) as formas funcionais dos modelos das médias não são todas iguais ou muito parecidas (conjuntos diferentes de efeitos para as respostas);
- b) dentre a maioria dos pares de equações, os resíduos têm correlações com valores estimados com magnitude superior a 0,40 e
- c) há modelo de variância estatisticamente significativo.

Se as condições (a) e (b) forem satisfeitas, recomenda-se refazer as estimativas sob a abordagem de sistema de equações de regressão aparentemente não relacionadas (SUR), conforme discutido na seção 2.3.2. Se, além disso, a condição (c) estiver presente, sugere-se aplicar o algoritmo desenvolvido no capítulo 4, que trata de uma extensão do método SUR para as situações em que as variâncias são funções dos fatores do processo.

Uma abordagem alternativa, que pode ser feita quando se supõe, previamente, que as respostas são correlacionadas, é identificar os efeitos significativos a partir de testes multivariados. Isto resulta num sistema de equações com a mesma matriz de planejamento. Neste tipo de sistema, a estimação univariada dos coeficientes é eficiente (ver seção 2.3.1).

A figura 5.2 ilustra as principais decisões e os procedimentos na construção de modelos com múltiplas respostas.

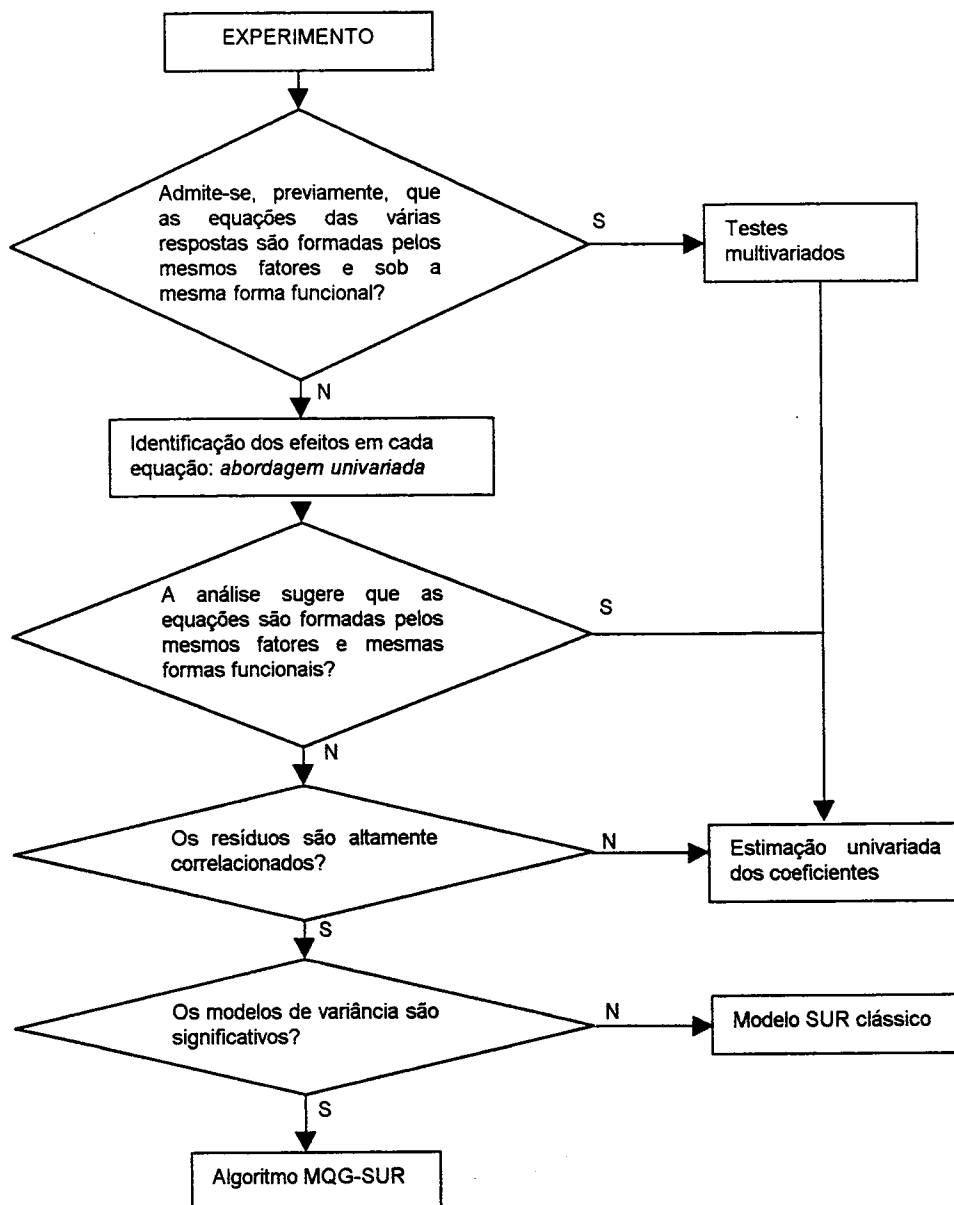


Fig. 5.2 – Estratégias para a construção de modelos com múltiplas respostas.

5.2 - ESTUDO DE CASO Nº 1: MOLDAGEM POR INJEÇÃO.

A presente aplicação foi extraída de Montgomery (1997, p. 391) e refere-se a um estudo experimental realizado com o intuito de se tentar reduzir o problema do encolhimento de determinado material moldado por injeção. Como a qualidade é prejudicada pelo encolhimento, buscaram-se meios que levem a um produto moldado com menor encolhimento possível.

Um grupo responsável pela qualidade decidiu conduzir um projeto de experimento estatisticamente planejado na tentativa de obter melhoria da qualidade sem aumentar os custos da produção. O grupo julgou que seis fatores controláveis do processo poderiam alterar o nível de encolhimento do material moldado. Estes fatores foram: a velocidade do parafuso de injeção (x_1), a temperatura do molde (x_2), o tempo em que o material é mantido no molde (x_3), o tempo de ciclo (x_4), o tamanho da entrada (x_5) e a pressão mantida no molde (x_6). Os fatores foram ensaiados a dois níveis segundo um experimento fatorial do tipo 2^k , com dois fracionamentos (2^{6-2}). A resposta do experimento foi o encolhimento do material moldado e, portanto, uma resposta do tipo *menor-é-melhor*. O projeto experimental e os resultados são apresentados na tabela 5.1.

Tabela 5.1 - Projeto experimental e os dados (grau de encolhimento) do experimento de moldagem por injeção.

ensaio	x_1	x_2	x_3	x_4	x_5	x_6	encolhimento observado
1	-	-	-	-	-	-	6
2	+	-	-	-	+	-	10
3	-	+	-	-	+	+	32
4	+	+	-	-	-	+	60
5	-	-	+	-	+	+	4
6	+	-	+	-	-	+	15
7	-	+	+	-	-	-	26
8	+	+	+	-	+	-	60
9	-	-	-	+	-	+	8
10	+	-	-	+	+	+	12
11	-	+	-	+	+	-	34
12	+	+	-	+	-	-	60
13	-	-	+	+	+	-	16
14	+	-	+	+	-	-	5
15	-	+	+	+	-	+	37
16	+	+	+	+	+	+	52

Fonte: Montgomery (1997, p. 392).

Nota. O sinal de negativo indica o nível codificado -1 e o sinal positivo representa +1.

5.2.1 - Identificação dos efeitos da média do processo

Para a identificação dos principais efeitos que alteram a média da resposta, construiu-se um gráfico de probabilidade normal dos efeitos (figura 5.3), considerando os seis efeitos principais e as nove interações possíveis de serem estimadas. Observa-se que o tipo de experimento em estudo impossibilita a utilização de procedimentos estatísticos mais formais.

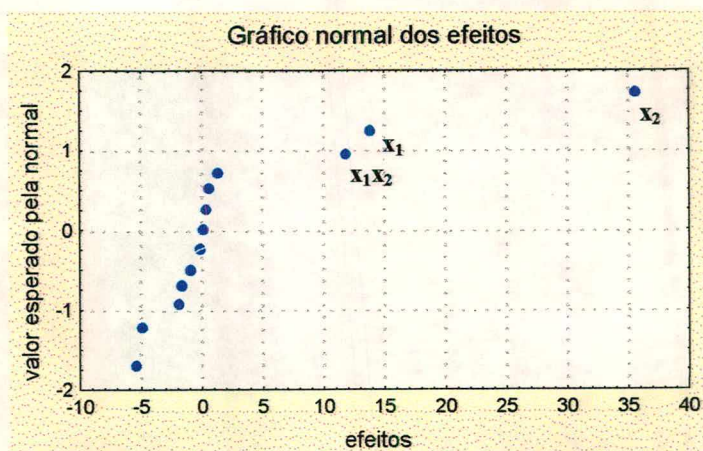


Fig. 5.3 - Gráfico normal dos efeitos da média do processo.

A figura 5.3 aponta claramente três efeitos separados da linha fictícia onde se agregam os efeitos casuais. Utilizando-se a técnica dos mínimos quadrados ordinários, obtém-se a seguinte equação para a média do processo:

$$\hat{\mu}_y = 27,31 + 6,94x_1 + 17,81x_2 + 5,94x_1x_2 \quad (5.6)$$

O modelo (5.6) é altamente significativo, com $F_0(3, 12) = 103$ ($p_F = 0,000$), $R^2 = 0,963$ e erros padrão das estimativas iguais a 1,14 e, portanto, bastante inferior às estimativas dos coeficientes.⁶ A figura 5.4 apresenta uma análise dos resíduos para se verificar a adequação do modelo (5.6).

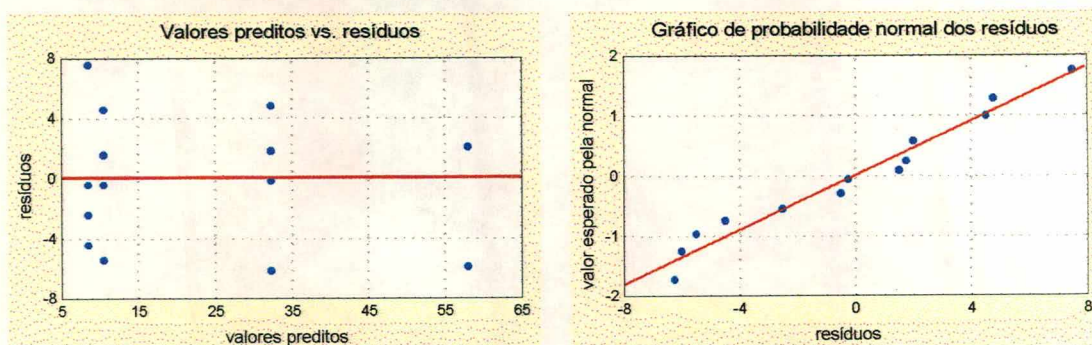


Fig. 5.4 - Análise dos resíduos do modelo da média.

Conforme sugerem os gráficos da figura 5.4, não há razão para se duvidar da adequação do modelo ajustado para a média.

5.2.2 - Identificação dos efeitos da variância do processo

Como o experimento não é replicado, os efeitos de dispersão só podem ser estudados a partir dos resíduos do modelo da média. Montgomery (1997) utiliza-se

⁶ Como o experimento é do tipo não replicado, as estatísticas são calculadas na suposição de que os efeitos que não aparecem na equação são variações casuais.

do método de Box e Meyer (1986a), conforme descrito na seção 2.2.1. A tabela 5.2 mostra os principais cálculos desta estatística, considerando todos os possíveis efeitos. Os confundimentos entre interações de segunda ordem também são mostrados na tabela 5.2 (cabeçalho). A figura 5.5 apresenta os valores da estatística F^* de Box e Meyer num gráfico de probabilidade normal.

Tabela 5.2 - Resíduos do modelo da média e o esquema de cálculo da estatística F^* para todos os efeitos possíveis.

#	x_1	x_2	x_1x_2 = x_3x_5	x_3	x_1x_3 = x_2x_5	x_1x_5 = x_2x_3	x_5	x_4	x_1x_4 = x_5x_6	x_2x_4 = x_3x_5	$x_1x_2x_4$ = x_3x_5	x_2x_6 = x_3x_5	$x_1x_3x_4$	x_6	x_1x_6 = x_4x_5	resíduos
1	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+	-2,50
2	+	-	-	-	-	+	+	-	-	+	+	+	+	-	-	-0,50
3	-	+	-	-	+	-	+	-	+	-	+	-	-	+	-	-0,25
4	+	+	+	-	-	-	-	-	-	-	-	+	+	+	+	2,00
5	-	-	+	+	-	-	+	-	+	+	-	-	+	+	-	-4,50
6	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+	4,50
7	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+	-6,25
8	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	2,00
9	-	-	+	-	+	+	-	+	-	-	+	-	+	+	-	-0,50
10	+	-	-	-	-	+	+	+	+	-	-	-	-	+	+	1,50
11	-	+	-	-	+	-	+	+	-	+	-	-	+	-	+	1,75
12	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-	2,00
13	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	7,50
14	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	-5,50
15	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	4,75
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-6,00
$S(i^+)$	3,80	4,01	4,33	5,70	3,68	3,85	4,17	4,64	3,39	4,01	4,72	4,71	3,50	3,88	4,87	
$S(i^-)$	4,60	4,41	4,10	1,63	4,53	4,33	4,25	3,59	2,75	4,41	3,51	3,65	3,12	4,52	3,40	
F_i^*	-0,38	-0,19	0,11	2,50	-0,42	-0,23	-0,04	0,51	0,42	-0,19	0,59	0,51	0,23	-0,31	0,72	

Fonte: Montgomery (1997, p. 396).

Síglas: $S(i^+)$ é o desvio padrão amostral dos resíduos com níveis +1 no i -ésimo efeito ($i = 1, \dots, 15$);
 $S(i^-)$ é o desvio padrão amostral dos resíduos com níveis -1 no i -ésimo efeito e
 F_i^* é a estatística de Box e Meyer (1986a) no i -ésimo efeito.

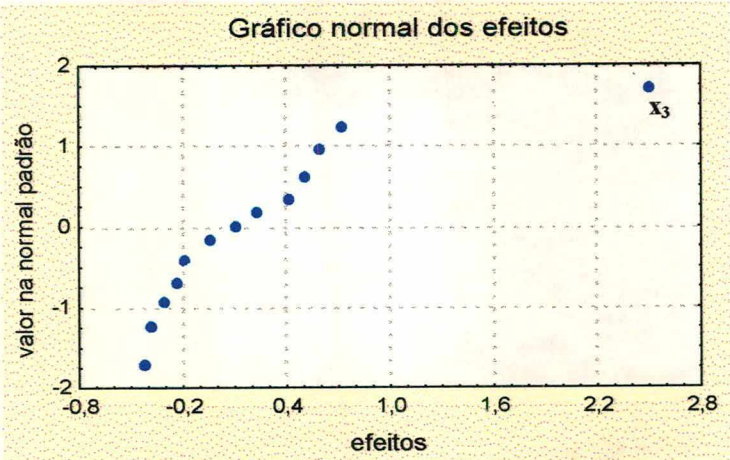


Fig. 5.5 - Gráfico normal dos efeitos de dispersão - estatística F_i^* de Box e Meyer

Pela figura 5.5, fica evidente apenas um efeito de dispersão, representado por x_3 . No nível mais baixo de x_3 (tempo em que o material é mantido no molde), tem-se menor variabilidade no grau de encolhimento, pois, conforme mostra a tabela 5.2, $S(\bar{i}^+) = 5,50$ e $S(\bar{i}^-) = 1,63$.

Refez-se a análise de identificação dos efeitos de dispersão utilizando-se o logaritmo dos resíduos quadráticos, $\log(\hat{e}_i^2)$, no lugar da estatística de Box e Meyer. O mesmo foi feito com a estatística $\log(\hat{e}_i^2 + \bar{e}^2)$, proposta neste trabalho. A tabela 5.3 mostra os valores destas estatísticas e as figuras 5.6 e 5.7 apresentam estes valores em gráficos de probabilidade normal.

Tabela 5.3 - Resíduos do modelo da média e transformações para a análise de identificação dos efeitos de dispersão.

#	resíduos \hat{e}_i	resíduos quadráticos \hat{e}_i^2	transformação log. padrão $\log(\hat{e}_i^2)$	transformação log. modificada $\log(\hat{e}_i^2 + \bar{e}^2)$
1	-2,50	6,25	1,83	3,08
2	-0,50	0,25	-1,39	2,76
3	-0,25	0,06	-2,77	2,75
4	2,00	4,00	1,39	2,97
5	-4,50	20,25	3,01	3,58
6	4,50	20,25	3,01	3,58
7	-6,25	39,06	3,67	4,00
8	2,00	4,00	1,39	2,97
9	-0,50	0,25	-1,39	2,76
10	1,50	2,25	0,81	2,88
11	1,75	3,06	1,12	2,92
12	2,00	4,00	1,39	2,97
13	7,50	56,25	4,03	4,27
14	-5,50	30,25	3,41	3,82
15	4,75	22,56	3,12	3,64
16	-6,00	36,00	3,58	3,94
média: $\bar{\hat{e}_i^2} = 15,55$				

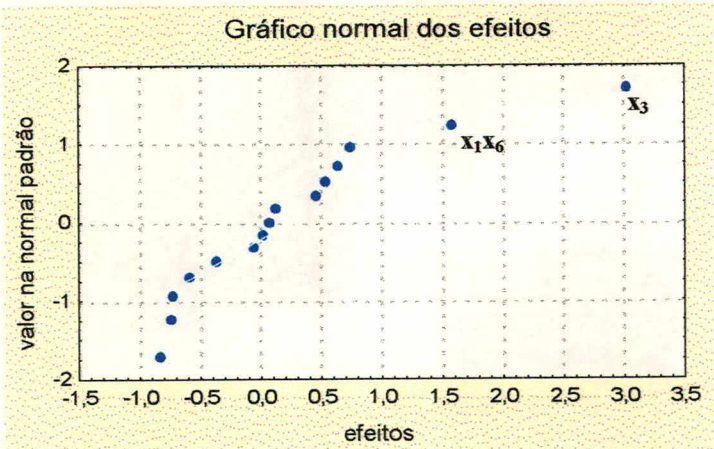


Fig. 5.6 - Gráfico normal dos efeitos de dispersão - estatística $\log(\hat{e}_i^2)$.

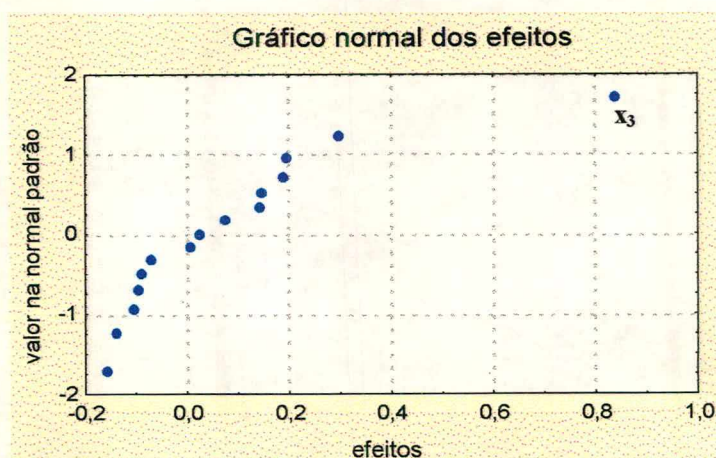


Fig. 5.7 - Gráfico normal dos efeitos de dispersão - estatística $\log(\hat{\epsilon}_i^2 + \bar{\epsilon}^2)$.

Na figura 5.6, quando se utiliza a transformação logarítmica padrão, o gráfico aponta para a possibilidade de haver também a interação x_1x_6 significativa, além do efeito principal de x_3 . Já o gráfico da figura 5.7 (transformação logarítmica modificada) apresenta-se muito parecido com o gráfico da estatística de Box e Meyer (figura 5.5), indicando apenas o efeito principal de x_3 como efeito de dispersão significativo. Embora não se conheça a realidade, a presença da interação x_1x_6 (ou de x_4x_5 , devido ao confundimento), na figura 5.6, é estranha, pois nenhum desses fatores tem efeito principal significativo, contrariando o princípio da hereditariedade, descrito por Box e Meyer (1986b) e Hamada e Wu (1992).

Ressalta-se que quando existe apenas um efeito real, a estatística de Box e Meyer é bastante eficiente. Nair e Pregibon (1988) mostram que, neste caso, o método de Box e Meyer é mais eficiente do que o método do logaritmo dos resíduos quadráticos ou do que o método da máxima verossimilhança. Assim, o fato da análise com a estatística aqui proposta gerar as mesmas conclusões do que a análise baseada na estatística de Box e Meyer é bastante interessante. Além disso, a estatística proposta pode ser usada para qualquer tipo de experimento, enquanto a de Box e Meyer é específica para experimentos fatoriais do tipo 2^k e, mesmo assim, funciona mal quando houver mais do que um efeito forte de dispersão, conforme mostram Nair e Pregibon (1988).

5.2.3 - Construção dos modelos da média e da variância

Myers e Montgomery (1995) argumentam que, em experimentos não replicados e com número reduzido de observações, como este em estudo, a construção de um modelo para a variância não é, necessariamente, uma boa prática, pois não se tem informação suficiente da variabilidade. No exemplo em questão, as condições ótimas são facilmente identificadas com a análise feita previamente: níveis mais baixos (-1) de x_1 e x_2 para garantir, em média, pequeno encolhimento (ver equação 5.6), e nível mais baixo de x_3 (-1) para garantir menor variabilidade do processo.

Com o objetivo de ilustrar os procedimentos propostos, segue a construção dos modelos. Adotam-se os dois níveis do fator de dispersão x_3 e consideram-se todos os outros efeitos que não se apresentaram significativos na análise de dispersão como se fossem *pseudo-replicações* e procede-se a análise aplicando-se MQO sobre o logaritmo dos resíduos quadráticos agregados.⁷ A equação resultante, após a transformação inversa, é apresentada a seguir:

$$\hat{\sigma}_y^2 = \exp\{3,03 + 1,22x_3\} \quad (5.7)$$

Na equação (5.7), o termo constante do argumento da função exponencial foi substituído pela estimativa do erro quadrático médio para evitar o vício devido a transformação logarítmica. Observa-se que devido à agregação dos fatores não significativos, a equação foi estimada com apenas dois pontos, impossibilitando o cálculo de qualquer estatística para a avaliação do modelo. A equação da variância pode ser usada para gerar pesos para estimar os parâmetros do modelo da média por MQG. Novos resíduos são calculados e novo modelo da variância pode ser construído. Após duas iterações, chega-se aos seguintes modelos:⁸

$$\hat{\mu}_y = 27,72 + 7,68x_1 + 18,67x_2 + 5,77x_1x_2 \quad (5.8)$$

$$\hat{\sigma}_y^2 = \exp\{2,88 + 1,57x_3\} \quad (5.9)$$

⁷ Para esta análise, é recomendável aplicar a correção de perda de graus de liberdade apresentada na seção 2.2.4. Ressalta-se que em experimentos do tipo 2^k a correção é a mesma para todos os resíduos.

⁸ O programa computacional encontra-se no apêndice.

Pela mesma razão citada anteriormente, o termo constante no argumento da função exponencial de (5.9) foi substituído pela estimativa do erro quadrático médio. As equações (5.8) e (5.9) formam as equações da média e da variância do processo, respectivamente.

5.3 - ESTUDO DE CASO Nº 2: O EXEMPLO DA CATAPULTA

Luner (1994) ilustra o uso do planejamento de experimentos num processo de melhoria da qualidade, utilizando-se de uma pequena catapulta construída com propósitos didáticos (ver fig. 5.8).

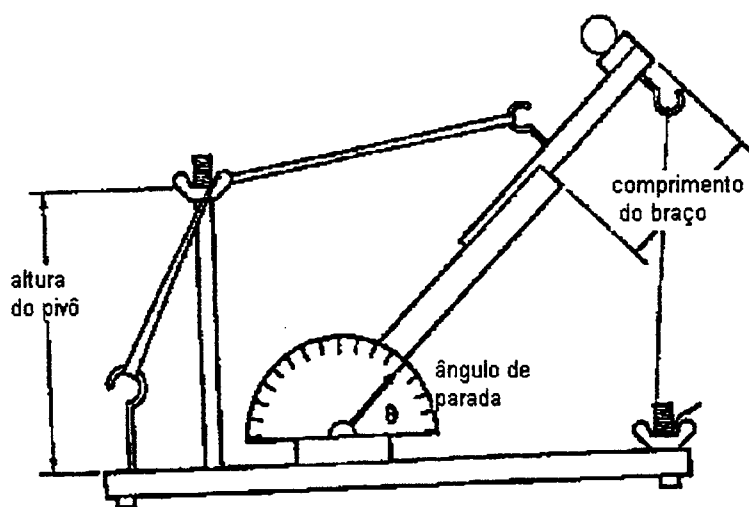


Fig. 5.8 - Esboço da catapulta com indicação dos fatores estudados no experimento.

Na definição do problema, o autor especifica que a catapulta deve arremessar projéteis a uma distância de 80 polegadas com alto grau de precisão.⁹ Além disso, deve ser robusta quanto a diferentes tipos de projéteis e outros possíveis ruídos. Com base no conhecimento do processo da catapulta, foram selecionados três fatores considerados críticos para a sua finalidade. Estes fatores são apresentados na tabela 5.4.

Tabela 5.4 - Fatores controláveis do processo usados no estudo experimental.

Fator	Amplitude máxima
comprimento do braço (x_1)	0,32 - 3,68 polegadas
ângulo de parada (x_2)	30 - 90°
altura do pivô (x_3)	2,5 - 5,5 polegadas

⁹ Optou-se por manter a unidade de medida do artigo original (1 polegada = 25,40 mm).

Considerou-se, também, que a catapulta deveria ser pouco sensível (robusta) a variações de projéteis, donde utilizou-se um fator de ruído, z , com três níveis (três tipos de bolas usadas como projéteis). O experimento foi realizado sob um projeto central composto (CCD), constituído por 8 pontos fatoriais, 6 axiais e 6 centrais. Em cada ponto experimental deste projeto, foram realizadas três observações, relativas aos diferentes níveis do fator de ruído z . A ordem dos ensaios foi aleatória. A resposta, y , foi definida como a distância, em polegadas, entre a base da catapulta e o local onde o projétil atinge o chão. A tabela 5.5 mostra os resultados do experimento.

Tabela 5.5 - Projeto experimental e resultados do experimento da catapulta.

#	fatores controláveis (níveis codificados)			níveis do fator de ruído e a resposta		
	x_1	x_2	x_3	$z = 1$	$z = 2$	$z = 3$
1	-1	-1	-1	39	34	42
2	-1	-1	1	80	71	91
3	-1	1	-1	52	44	45
4	-1	1	1	97	68	60
5	1	-1	-1	60	53	68
6	1	-1	1	113	104	127
7	1	1	-1	78	64	65
8	1	1	1	130	79	75
9	-1,682	0	0	59	51	60
10	1,682	0	0	115	102	117
11	0	-1,682	0	50	43	57
12	0	1,682	0	88	49	43
13	0	0	-1,682	54	50	60
14	0	0	1,682	122	109	119
15	0	0	0	87	78	89
16	0	0	0	86	79	85
17	0	0	0	88	81	87
18	0	0	0	89	82	87
19	0	0	0	86	79	88
20	0	0	0	88	79	90

Fonte. Luner (1994).

5.3.1 - Ajuste inicial do modelo da média

Luner (1994) construiu o modelo da média do processo com base nas médias amostrais, calculadas sobre os níveis do fator de ruído. Isto, segundo o autor, elimina a variação provocada pelo ruído z , a qual não precisa ser considerada ao avaliar a média do processo. Neste trabalho, contudo, o modelo será estimado diretamente das observações, como sugerido, por exemplo, por Myers e Montgomery (1995).

Especificou-se um modelo quadrático completo e fizeram-se, inicialmente, as estimações dos parâmetros pelo método dos mínimos quadrados ordinários (MQO).

A tabela 5.6 apresenta as estimativas dos coeficientes e as tabelas 5.7 e 5.8 apresentam estatísticas que permitem uma análise global do modelo ajustado. Estes resultados foram obtidos pelo procedimento *rsreg* do SAS.

Tabela 5.6 - Construção de um modelo para a média do processo da catapulta: ajuste de um modelo quadrático completo por MQO.

termo	estimativa do coeficiente	erro padrão	t_0	p_t
constante	84,95	2,52	33,66	0,000
x_1	13,88	1,67	8,29	0,000
x_2	0,62	1,67	0,37	0,712
x_3	18,64	1,67	11,13	0,000
x_1^2	-0,70	1,63	-0,42	0,670
x_2^2	-10,95	1,63	-6,72	0,000
x_3^2	-0,11	1,63	-0,07	0,947
x_1x_2	-1,79	2,19	-0,82	0,417
x_1x_3	1,21	2,19	0,55	0,583
x_2x_3	-5,38	2,19	-2,46	0,018
$F_0(9, 50) = 27,27$ ($p_F = 0,000$) $R^2 = 0,831$				
Siglas: t_0 valor da estatística t sob a hipótese de que o coeficiente é nulo; p_t nível descritivo de significância (teste t bilateral); R^2 coeficiente de determinação e $F_0(., .)$ valor da estatística F sob a hipótese de que todos os coeficientes são nulos, com indicação dos graus de liberdade.				

Tabela 5.7 - Avaliação do modelo especificado: verificação de falta de ajuste.

resíduo	graus de liberdade	soma de quadrados	F_0	p_F
falta de ajuste	5	433	0,734	0,602
erro puro	45	5309		
total	50	5742		
Siglas: F_0 valor da estatística F sob H_0 e p_F nível descritivo de significância.				

Tabela 5.8 - Avaliação do modelo especificado: decomposição

regressão	graus de liberdade	soma de quadrados	F_0	p_F
termos com x_1	4	8030	17,48	0,000
termos com x_2	4	5973	13,00	0,000
termos com x_3	4	14969	32,58	0,000
linear	3	22153	64,29	0,000
quadrática	3	5231	15,18	0,000
interações	3	805	2,34	0,085
completa	9	28190	27,27	0,000
Siglas: F_0 valor da estatística F sob a hipótese de que o conjunto de coeficientes é nulo (na regressão linear assume-se que a constante já faz parte do modelo e nas regressões quadrática e de interações assume-se, também, que os termos lineares estejam incluídos no modelo) e p_F nível descritivo de significância do teste F.				

Conforme as estatísticas da tabela 5.6, o modelo apresentou alto coeficiente de determinação ($R^2 = 0,831$) e a maioria dos coeficientes mostraram-se significativos. A tabela 5.7 mostra que o modelo de segunda ordem é suficiente, pois não há evidência de falta de ajuste. A tabela 5.8 indica que os termos lineares e quadráticos puros são fundamentais na especificação deste modelo, enquanto as interações poderiam ser excluídas caso a aplicação exigisse um modelo bastante parcimonioso. Além disso, esta tabela também mostra que todos os fatores em estudo são relevantes, sejam através dos efeitos lineares, quadráticos ou de interação com outros fatores.

Com a presente análise e considerando que um modelo quadrático completo pode ser interpretado como uma aproximação de segunda ordem da verdadeira superfície de resposta, optou-se por considerar todos os termos no modelo. A figura 5.9 apresenta a análise de resíduos do modelo apresentado na tabela 5.6. Os gráficos foram construídos com o programa computacional *Statistica*, versão 5.0.

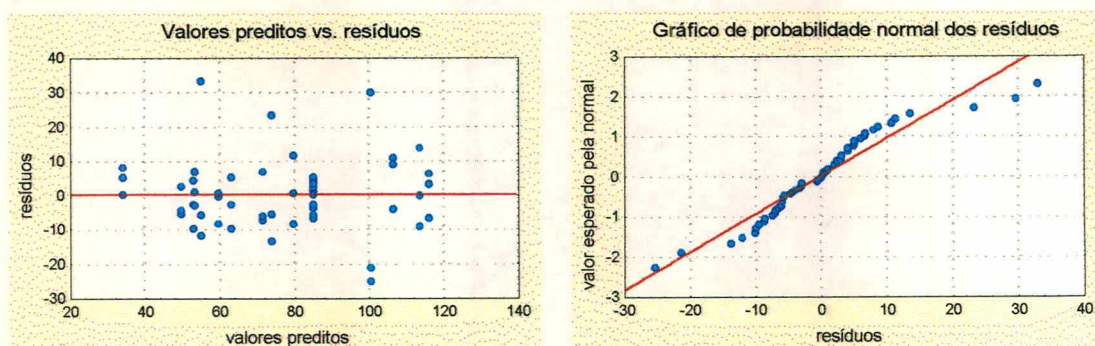


Fig. 5.9 - Análise de resíduos do modelo da média. Estimação por MQO.

Os gráficos da figura 5.9 mostram a presença de alguns pontos discrepantes. Uma análise destes pontos revelam que três deles pertencem a um mesmo ponto experimental, sendo duas observações com resíduos negativos e uma com resíduo positivo. Assim, a exclusão destes pontos, além de deixar os dados desbalanceados, pode eliminar informações relevantes a cerca da variância do processo. Outro fato que levou a decisão de manter tais pontos é que a estimação por MQG, a ser feita

posteriormente, reduz a influência de pontos discrepantes na estimação dos parâmetros do modelo.¹⁰

5.3.2 - Ajuste inicial do modelo da variância

Como o modelo da média ajustou-se adequadamente aos dados, com o teste F de falta de ajuste acusando nível descritivo de significância superior a 0,50 ($p_F = 0,602$), podem-se usar os resíduos quadráticos como as medidas primárias para o estabelecimento do modelo da variância do processo. Observa-se, também, que a medida agregada de variância, definida pela expressão (3.33), atribui peso nulo para as variâncias amostrais e peso um para os resíduos quadráticos médios. Os resíduos quadráticos médios, \hat{e}_i^2 ($i = 1, 2, \dots, 20$), foram calculados ao longo dos pontos experimentais, sendo que no ponto central calculou-se um resíduo quadrático médio para cada trinca de ensaios. A tabela 5.9 apresenta os valores calculados dos \hat{e}_i^2 ($i = 1, 2, \dots, 20$), considerando a correção da perda dos graus de liberdade no modelo da média, conforme a expressão (2.31).

Tabela 5.9 - Resíduos quadráticos médios calculados sobre os níveis do ruído.

#	\hat{e}_i^2	#	\hat{e}_i^2	#	\hat{e}_i^2	#	\hat{e}_i^2
1	37,305	6	117,149	11	51,688	16	12,875
2	87,318	7	59,915	12	530,174	17	10,273
3	25,391	8	849,036	13	23,574	18	10,347
4	327,219	9	31,284	14	39,302	19	16,162
5	57,601	10	86,967	15	24,316	20	24,779

A equação de regressão foi feita em termos do $\log(\hat{e}_i^2 + c\bar{e}^2)$. O valor de c foi calculado pela expressão (3.28), com $A = 0$, $D = 1$ e $F = 0$, obtendo-se $c = 0,08$. Os valores de A e D baseiam-se no projeto de experimentos em questão e o valor F foi arbitrado (corresponde ao coeficiente de maior magnitude no modelo da variância, assumido, inicialmente, com valor igual a 0,7).

¹⁰ Verificou-se, também, que apesar dos pontos serem discrepantes, eles não são muito influentes na estimação dos parâmetros. O cálculo da estatística D de Cook revela valores sempre abaixo de 0,31. Myers e Montgomery (1995, p. 50) consideram críticas as observações com valores de D acima de um.

A tabela 5.10 apresenta as estimativas dos parâmetros de um modelo quadrático completo para a variância.

Tabela 5.10 - Construção de um modelo para a variância do processo da catapulta: ajuste por MQO de um modelo quadrático completo com transformação logarítmica modificada.

termo	estimativa do coeficiente	erro padrão	t ₀	p _t
constante ¹	4,80	-	-	-
x ₁	0,27	0,13	2,08	0,065
x ₂	0,48	0,13	3,69	0,004
x ₃	0,50	0,13	3,81	0,003
x ₁ ²	0,39	0,13	3,05	0,012
x ₂ ²	0,76	0,13	6,01	0,000
x ₃ ²	0,23	0,13	1,81	0,100
x ₁ x ₂	0,12	0,17	0,73	0,482
x ₁ x ₃	0,02	0,17	0,12	0,909
x ₂ x ₃	0,43	0,17	2,51	0,031
F ₀ (9, 10) = 9,16 (p _F = 0,001)			R ² = 0,892	

¹ Neste termo, a estimativa de MQO foi substituída pelo logaritmo da estimativa do erro quadrático médio, $\log(\hat{\epsilon})$.

As estatísticas da tabela 5.10 mostram um modelo significativo para a variância, com a maioria dos coeficientes significativamente diferentes de zero. Testes do tipo F, assim como feitos no modelo da média (tabela 5.8) poderiam ser realizados para se tentar um modelo mais simples para a variância, mas optou-se por considerar o modelo quadrático completo. Poder-se-iam, também, usar as estimativas iniciais dos coeficientes para reavaliar o valor de *c* da transformação logarítmica utilizada, mas verifica-se que o coeficiente de maior magnitude tem valor próximo de 0,7, como foi inicialmente arbitrado. Logo, neste caso em particular, não é necessário alterar a transformação. A figura 5.10 apresenta a análise de resíduos deste modelo.

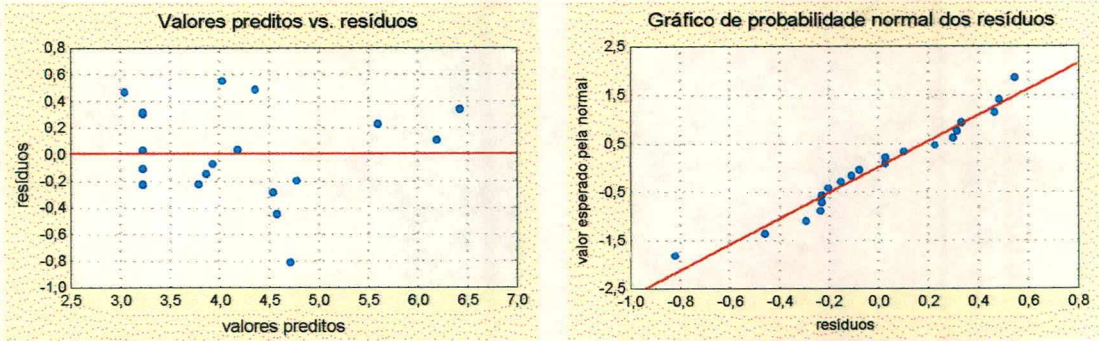


Fig. 5.10 - Análise de resíduos do modelo da variância. Estimação por MQO após transformação logarítmica modificada.

A distribuição dos resíduos (figura 5.10) mostra-se razoável. O resíduo com maior magnitude não se encontra numa das extremidades do eixo dos valores preditos. Logo, não há motivos para excluí-lo.

5.3.3 - Estimação conjunta dos parâmetros dos modelos da média e da variância

Com os modelos da média e da variância especificados como superfícies de resposta de segundo grau, procedeu-se a estimação conjunta dos parâmetros utilizando-se MQG para o modelo da média e MQO com transformação logarítmica para o modelo da variância, conforme discutido na seção 5.1.2. As tabelas 5.11 e 5.12 apresentam os resultados após duas iterações e as figuras 5.11 e 5.12 mostram a análise dos resíduos destes modelos, em termos das variáveis transformadas. O programa computacional encontra-se no apêndice.

Tabela 5.11 - Modelo final para a média do processo da catapulta.

termo	estimativa do coeficiente	erro padrão	t_0	p_t
constante	84,88	1,08	78,48	0,000
x_1	14,81	1,32	11,20	0,000
x_2	0,19	1,77	0,10	0,917
x_3	19,04	1,42	13,45	0,000
x_1^2	-0,67	1,03	-0,65	0,520
x_2^2	-11,57	1,45	-7,99	0,000
x_3^2	0,58	1,02	0,57	0,573
x_1x_2	0,11	1,68	0,07	0,948
x_1x_3	2,89	1,66	1,74	0,088
x_2x_3	-4,66	1,86	-2,51	0,015
$F_0(9, 50) = 182,67$ ($p_F = 0,000$)			$R^2 = 0,970$	

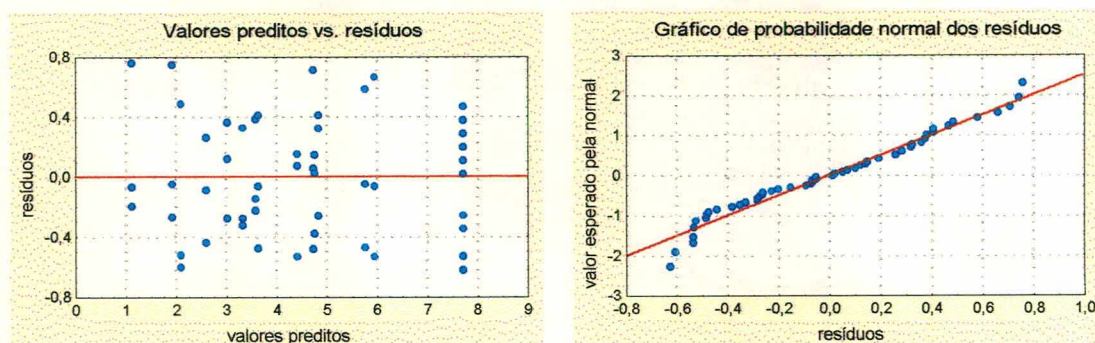


Fig. 5.11 - Análise de resíduos do modelo final da média. Estimação por MQG.

Tabela 5.12 - Modelo final para a variância do processo da catapulta. Estatísticas em termos do modelo exponencial.

termo	estimativa do coeficiente	erro padrão	t_0	p_t
constante ¹	4,80	-	-	-
x_1	0,32	0,14	2,21	0,051
x_2	0,51	0,14	3,59	0,005
x_3	0,58	0,14	4,04	0,002
x_1^2	0,31	0,14	2,25	0,048
x_2^2	0,72	0,14	5,21	0,000
x_3^2	0,24	0,14	1,70	0,120
x_1x_2	0,09	0,19	0,50	0,627
x_1x_3	-0,08	0,19	-0,43	0,676
x_2x_3	0,36	0,19	1,96	0,079
$F_0(9, 10) = 7,74 \quad (p_F = 0,002)$			$R^2 = 0,874$	

¹ Neste termo, a estimativa de MQO foi substituída pelo logaritmo da estimativa do erro quadrático médio, $\log(\hat{\epsilon})$.

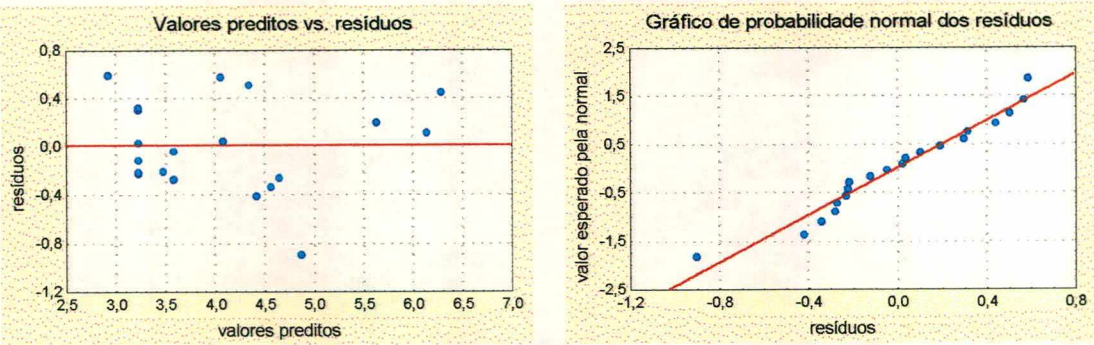


Fig. 5.12 - Análise de resíduos do modelo final da variância, construído com os resíduos do modelo da média, o qual foi ajustado por MQG.

Comparando o modelo final da média (tabela 5.11) com o modelo inicial (tabela 5.6), verifica-se que os coeficientes não tiveram grandes alterações, contudo a qualidade do ajuste do modelo parece bem melhor no modelo final, obtido por MQG. A razão F_0 aumentou de 27 para 182 e o coeficiente R^2 aumentou de 0,83 para 0,97. Ressalta-se, porém, que a comparação entre os dois modelos pelas estatísticas F_0 e R^2 deve ser feita com cautela, pois são calculadas sob diferentes métricas. Em termos gráficos, a figura 5.11 mostra uma distribuição dos resíduos bem melhor comportada do que a figura 5.9. Observa-se que os resíduos apresentados na figura 5.11 foram calculados com os pesos usados no MQG. Com estas informações, pode-se ter melhor garantia da equação produzida pelo método MQG do que a gerada pelo método MQO.

Os modelos da variância, construídos com os resíduos do modelo da média ajustado por MQO (tabela 5.10) e por MQG (tabela 5.12), apresentaram-se bem

parecidos, tanto em termos dos coeficientes estimados quanto nos gráficos dos resíduos (figuras 5.10 e 5.12). Os valores das estatísticas F_0 e R^2 , contudo, ficaram levemente reduzidas no segundo caso. Considerando, porém, que o método MQG produziu um modelo da média aparentemente mais adequado do que o método MQO, parece razoável admitir que um modelo da variância baseado nos resíduos do MQG é mais confiável.

Em resumo, tem-se, então, as seguintes equações para representar a média e a variância do processo da catapulta:

$$\hat{\mu}_y = 84,88 + 14,81x_1 + 0,19x_2 + 19,04x_3 - 0,67x_1^2 - 11,57x_2^2 + 0,58x_3^2 + 0,11x_1x_2 + 2,89x_1x_3 - 4,66x_2x_3 \quad (5.10)$$

$$\hat{\sigma}_y^2 = \exp\{4,80 + 0,32x_1 + 0,51x_2 + 0,58x_3 + 0,31x_1^2 + 0,72x_2^2 + 0,24x_3^2 + 0,09x_1x_2 - 0,08x_1x_3 + 0,36x_2x_3\} \quad (5.11)$$

A tabela 5.13 mostra uma análise comparativa da qualidade do ajuste do modelo (5.11) com outros procedimentos bastante utilizados. Estes outros procedimentos baseiam-se nas variâncias amostrais, calculadas sobre os níveis do fator de ruído. Portanto, 20 variâncias amostrais calculadas sobre 3 observações (S_i^2 , $i = 1, 2, \dots, 20$), correspondentes a cada linha da tabela 5.6. Foram usadas três transformações nas variâncias amostrais: a logarítmica padrão, a raiz quadrada e a logarítmica modificada com $c = 0,08$. A transformação logarítmica padrão é a mais comum neste tipo de problema, a raiz quadrada foi utilizada por Luner (1994) e a última foi proposta neste trabalho.

Tabela 5.13 - Comparação de procedimentos para estabelecer modelos para a variância, considerando os dados do experimento da catapulta.

transformação	R^2	F_0
$\log(S_i^2)$	0,854	6,47
$\sqrt{S_i^2}$	0,857	6,68
$\log(S_i^2 + c\bar{S}^2)$	0,857	6,64
$\log(\hat{e}_i^2 + c\bar{\hat{e}}^2)^*$	0,892	9,16
$\log(\hat{e}_i^2 + c\bar{\hat{e}}^2)^{**}$	0,874	7,74

* \hat{e}_i representam resíduos gerados por MQO.

** \hat{e}_i representam resíduos gerados por MQG.

Comparando as três primeiras linhas da tabela 5.13, verifica-se melhor adequação das transformações $\sqrt{S_i^2}$ e $\log(S_i^2 + c\bar{S}^2)$. Quando se planeja um modelo para a variância propriamente dita, a transformação raiz quadrada cria alguns problemas, pois a sua transformação inversa (potência quadrática), realizada após o processo de estimação dos parâmetros, produz um modelo com termos do tipo $x_1^2x_2$ ou $x_1x_2x_3$, que são difíceis de serem interpretados na prática. Modelos para a variância propriamente dita são relevantes quando se busca a otimização através da função perda quadrática, bastante comum na engenharia da qualidade. Luner (1994) utilizou a metodologia da superfície de resposta dual, onde um modelo para o desvio padrão do processo é suficiente (ver seção 2.4).

A tabela 5.13 também mostra que as melhores adequações foram obtidas com modelos construídos a partir dos resíduos.¹¹ Ressalta-se, porém, que modelos baseados nos resíduos dependem da adequação do modelo da média do processo. Na abordagem desenvolvida neste trabalho, esta adequação é garantida com a medida agregada de variância, definida pela expressão (3.33).

5.4- ESTUDO DE CASO Nº 3: FORMULAÇÃO DE UMA MASSA CERÂMICA

Bernardin (1994) fez um estudo experimental, baseado na metodologia de Taguchi, para melhorar o processo de formulação de massa cerâmica para pavimento. A qualidade foi avaliada pela retração linear do corpo queimado (%), resistência mecânica (Mpa) e absorção de água (%). Foram considerados como fatores controláveis a temperatura de queima, os percentuais de filito, argila, talco e arenito e os tipos de argila, filito e talco. Desta forma, o estudo experimental foi constituído por três respostas (y_1 , y_2 e y_3) e oito fatores (x_1 , x_2 , ..., x_8), cujos níveis são apresentados na tabela 5.14.

¹¹ Na verdade, usou-se a medida agregada de variância, v_i ($i = 1, 2, \dots, 20$), definida na seção 3.4. Contudo, em todas as iterações do processo de estimação, a medida agregada teve peso um para os resíduos quadráticos médios e peso nulo para as variâncias amostrais.

Tabela 5.14 - Níveis dos fatores controláveis do processo da massa cerâmica.

fator	nível -1	nível 0	nível +1
x ₁ : temperatura de queima (°C)	1140	–	1160
x ₂ : percentual de filito (%)	43	46	49
x ₃ : percentual de argila (%)	22	25	28
x ₄ : percentual de talco (%)	9	12	15
x ₅ : percentual de arenito (%)	8	11	14
x ₆ : tipo de argila	PR 1	PR 1 / DL 34	DL 34
x ₇ : tipo de filito	ITAPEVA	ITAPEVA / LV	LV
x ₈ : tipo de talco	ITA 124	ITA 124 / G1	G1

O projeto experimental usado por Bernardin foi o L₁₈, com o fator x₁ fixado em dois níveis e os outros fatores em três níveis. O projeto L₁₈ é largamente usado na indústria e consiste de 18 pontos experimentais, combinando fatores com dois e três níveis (ver, por exemplo, Taguchi, 1987). Foram realizados cinco ensaios em cada uma das 18 combinações de níveis dos fatores. Contudo, 9 desses ensaios tiveram falta de informação em alguma das variáveis, resultando num total de 81 observações completas do vetor (y₁, y₂, y₃)'. Os dados deste experimento encontram-se no apêndice, juntamente com o programa computacional.

Caten (1995) fez uma série de críticas ao estudo de Bernardin (1994). As críticas citam a falta de respostas relevantes do processo, como a defloculação, a falta de um pré-planejamento para escolher melhor os fatores e os seus níveis e a própria deficiência da metodologia de Taguchi, utilizada na análise. Registra-se que, na análise de Bernardin (1994) e de Caten (1995), os resultados deste experimento foram agregados com os de outro experimento realizado posteriormente, o qual não está sendo considerado neste trabalho.

5.4.1 - Modelos iniciais para as médias do processo.

Seguindo Caten (1995), admitiu-se que todos os fatores são contínuos, em especial, os tipos de argila, filito e talco. Ou seja, considerou-se que estes tipos são misturas que podem ser feitas com quaisquer proporções dos ingredientes citados na tabela 5.14, sem restrições. Além disso, assumiu-se que o nível intermediário (0) corresponde à situação média entre os extremos (–1 e +1).¹²

¹² Sem estas suposições, o modelo poderia ser ajustado com a transformação de cada um destes três fatores em duas variáveis indicadoras (*dummy variables*), as quais poderiam ser analisadas em termos de seus efeitos principais e de interações com os outros fatores.

Na fase de identificação dos fatores, optou-se por considerar uma resposta de cada vez, ao contrário da abordagem de regressão multivariada (ou análise de variância multivariada), onde se verifica a significância dos efeitos em termos do vetor de respostas como um todo.

Com sete fatores ensaiados a três níveis e um ensaiado a dois níveis, ter-se-iam 44 parâmetros num modelo polinomial com efeitos lineares, de interações de segunda ordem e, para os fatores ensaiados a três níveis, efeitos quadráticos. Como o projeto experimental L_{18} tem ensaios em apenas 18 pontos da região experimental, torna-se inviável ter estimativas para todos estes parâmetros. Adotou-se, então, a abordagem sugerida por Hamada e Wu (1992), conforme discutido na seção 2.1.2. Esta abordagem baseia-se nos princípios de *esparsidade* e *hereditariedade*. O princípio da *esparsidade* considera que apenas uma pequena parte dos efeitos são relevantes no processo e o princípio da *hereditariedade* considera que não faz sentido físico uma interação do tipo x_1x_2 ser significativa, se nenhum dos efeitos principais (x_1 ou x_2) o forem.

Para cada resposta, verificou-se, primeiramente, a significância de cada um dos efeitos principais (lineares e quadráticos) e, depois, a significância dos termos das interações de primeira ordem, mas restringindo-se àquelas onde pelo menos um dos fatores teve efeito linear significativo na primeira etapa. Devido às restrições do planejamento L_{18} , os efeitos foram testados um de cada vez. Com as interações significativas incluídas no modelo, voltou-se a observar a significância dos efeitos principais. Tendo em vista o objetivo de obter modelos adequados para realizar previsões, levou-se em conta na seleção de cada termo, além do nível de significância do teste t, o incremento no coeficiente R^2 ajustado. Assim, alguns termos foram incluídos (ou excluídos) mesmo com o teste t acusando (não acusando) significância.

A tabela 5.15 mostra as estimativas de MQO dos coeficientes que se mostraram mais relevantes em cada equação de regressão da média. Os cálculos foram feitos em termos dos valores codificados dos fatores.

Tabela 5.15 - Efeitos identificados em cada equação de regressão da média, coeficientes estimados por MQO, erros padrão das estimativas e algumas estatísticas de ajuste dos modelos.

Efeito	Retração linear		Resistência mecânica		Absorção de água	
	coef.	erro padrão	coef.	erro padrão	coef.	erro padrão
intercepto	11,50	0,35	43,87	1,22	3,27	0,23
x ₂	-0,90	0,13	-4,18	0,84	0,87	0,09
x ₃	-0,86	0,13	-5,10	0,78	1,43	0,08
x ₅	-0,78	0,12	-5,11	0,82	1,00	0,07
x ₆	0,47	0,12	3,98	0,82	-1,74	0,09
x ₇	-	-	-	-	0,38	0,07
x ₂ ²	-1,23	0,21	-	-	1,76	0,14
x ₃ ²	0,91	0,21	-	-	-0,84	0,15
x ₄ ²	0,85	0,22	-	-	-1,70	0,13
x ₅ ²	-0,70	0,21	-	-	0,45	0,15
x ₆ ²	-0,48	0,21	-	-	-	-
x ₇ ²	-2,29	0,21	-3,93	1,38	1,98	0,13
x ₁ x ₂	-	-	-	-	-0,29	0,09
x ₂ x ₃	-	-	3,06	1,11	-0,40	0,12
x ₃ x ₆	-	-	-5,35	1,19	-	-
x ₄ x ₇	-	-	-	-	-0,51	0,16
x ₅ x ₆	0,53	0,18	-	-	-	-
Modelo	F ₀ (11,69)=39,4 (p=0,000) R ² =0,863		F ₀ (7,73)=20,8 (p=0,000) R ² =0,666		F ₀ (13,67)=166 (p=0,000) R ² =0,970	
Falta de ajuste	F ₀ (6,63)=0,877 (p=0,517)		F ₀ (10,63)=0,845 (p=0,587)		F ₀ (4,63)=1,461 (p=0,225)	

Verificam-se nos testes F, tabela 5.15, que todos os modelos são significativos e que não há evidência de falta de ajuste em nenhum deles. A figura 5.13 mostra uma análise dos resíduos de cada um destes modelos.

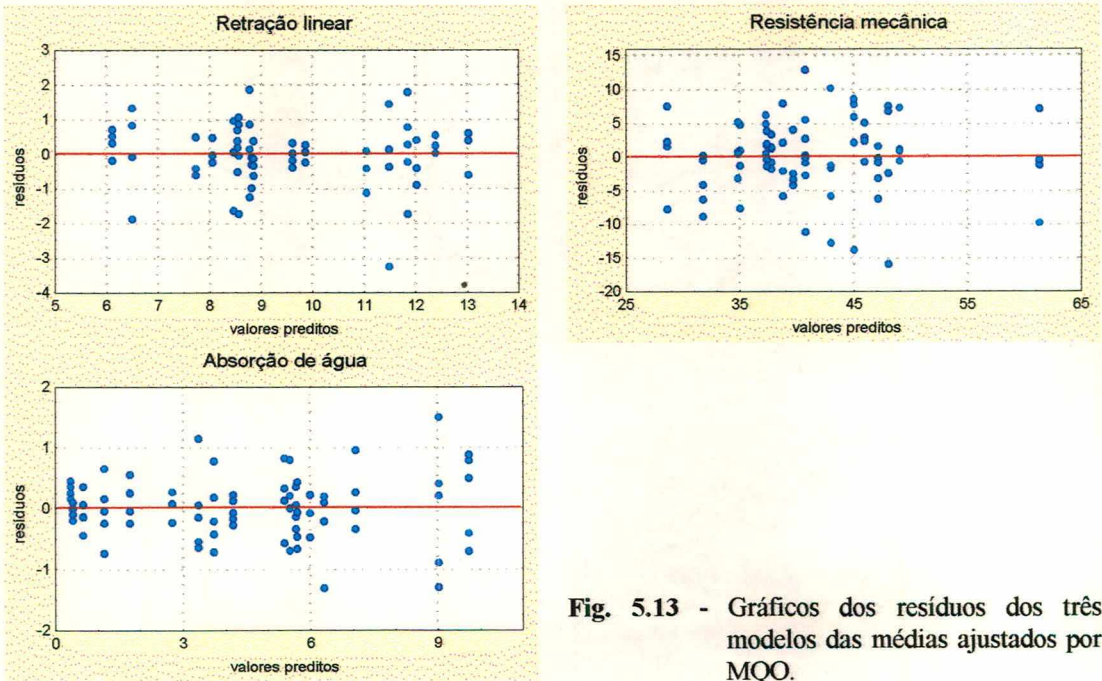


Fig. 5.13 - Gráficos dos resíduos dos três modelos das médias ajustados por MQO.

Analisando a figura 5.13, verifica-se que os resíduos das variáveis em estudo comportam-se de forma parecida com uma distribuição normal e não há evidência de algum padrão que sugira transformações. Somente em relação ao gráfico da retração linear, observa-se um resíduo ligeiramente discrepante dos demais, mas optou-se por não se fazer tratamento diferenciado para este dado, pelo menos nesta fase da análise.

5.4.2 - Modelos para as variâncias do processo.

Como o número de replicações do experimento em discussão não é muito pequeno (igual a cinco nas condições experimentais sem valores faltantes), partiu-se da informação fornecida pela variância amostral, calculada em cada condição experimental, para se construir os modelos das variâncias. Com os valores desta estatística, pode-se construir modelos para as variâncias, independentemente da qualidade do ajuste dos modelos das médias.

Numa análise preliminar, verificou-se o ajuste de equações de regressão em função de $\log(S_i^2)$ e em função de $\log(S_i^2 + c\bar{S}^2)$, com $c = 0,02$ ($i = 1, 2, \dots, 18$). Para as três respostas, verificaram-se melhores ajustes com a transformação logarítmica modificada. Neste estudo, o valor de c foi arbitrário, mas se fosse usada a equação (5.2), com $F = 0$ e $A = -0,37$, obter-se-ia um valor de c apenas ligeiramente maior do que o valor arbitrado.¹³

Utilizando-se, para cada resposta, $\log\{S_i^2 + (0,02)\bar{S}^2\}$ como variável dependente e seguindo a mesma abordagem aplicada nos modelos das médias, chegaram-se às equações de regressão apresentadas na tabela 5.16.

¹³ O valor de A foi obtido ajustando-se a equação para se usar o número de pontos experimentais no lugar do número de fatores.

Tabela 5.16 - Efeitos identificados em cada equação de regressão da variância, coeficientes estimados por MQO com transformação $\log\{S_i^2 + (0,02)\bar{S}^2\}$, erros padrão das estimativas e algumas estatísticas de ajuste dos modelos.

Efeito	Retração linear		Resistência mecânica		Absorção de água	
	coef.	erro padrão	coef.	erro padrão	coef.	erro padrão
intercepto ¹	-0,32	-	3,46	-	-1,26	-
x ₂	-	-	-0,30	0,11	-	-
x ₄	-0,31	0,21	-	-	-	-
x ₅	0,35	0,21	-	-	-	-
x ₆	-	-	-	-	-0,82	0,22
x ₇	0,94	0,21	0,62	0,11	0,76	0,19
x ₈	0,31	0,21	-	-	-	-
x ₂ ²	-	-	-	-	0,43	0,33
x ₅ ²	0,54	0,37	-	-	0,81	0,33
x ₆ ²	-	-	-1,15	0,19	-	-
x ₇ ²	0,66	0,37	-	-	-	-
x ₈ ²	-0,84	0,37	-1,20	0,19	-	-
x ₁ x ₅	-	-	-	-	-0,58	0,23
x ₅ x ₇	-	-	-	-	-0,47	0,26
Modelo	F ₀ (7, 10) = 5,34 (p _F = 0,009) R ² = 0,789		F ₀ (4, 13) = 28,6 (p _F = 0,000) R ² = 0,898		F ₀ (6, 11) = 7,14 (p _F = 0,003) R ² = 0,796	

¹ A estimativa de MQO do termo constante foi substituída pelo logaritmo da média das variâncias amostrais, $\log(\bar{S}^2)$.

Os resultados apresentados na tabela 5.16 mostram que existe influência significativa dos fatores sobre a variância do processo em todas as equações (ver os resultados dos testes F sobre os modelos). Realizou-se análise de resíduos para as três equações de regressão e não se verificou qualquer inadequação dos modelos ajustados (ver figura 5.14).

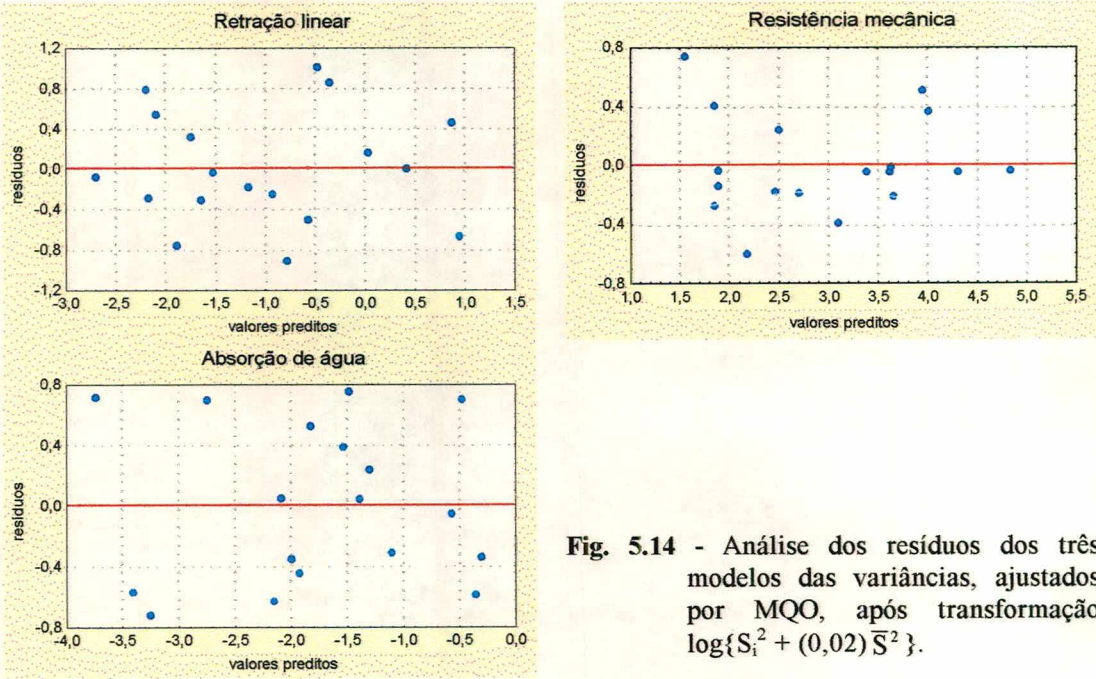


Fig. 5.14 - Análise dos resíduos dos três modelos das variâncias, ajustados por MQO, após transformação $\log\{S_i^2 + (0,02)\bar{S}^2\}$.

5.4.3 - Modelos univariados para as médias usando MQG.

Tendo em vista a existência de fatores que alteram a variância do processo, voltou-se a estimar os modelos das médias através da técnica MQG, tomando-se como peso o inverso das variâncias preditas. A tabela 5.17 apresenta os principais resultados.

Tabela 5.17 - Equações de regressão das médias, com estimativas por MQG, tratando cada resposta isoladamente.

Efeito	Retração linear		Resistência mecânica		Absorção de água	
	coef.	erro padrão	coef.	erro padrão	coef.	erro padrão
intercepto	11,62	0,26	45,28	0,80	3,17	0,22
x ₂	-0,77	0,11	-5,66	0,66	0,88	0,09
x ₃	-0,92	0,10	-4,87	0,60	1,40	0,07
x ₅	-0,82	0,10	-5,14	0,56	0,98	0,08
x ₆	0,40	0,10	4,55	0,50	-1,73	0,07
x ₇	-	-	-	-	0,35	0,08
x ₂ ²	-1,36	0,19	-	-	1,74	0,13
x ₃ ²	1,14	0,16	-	-	-0,86	0,14
x ₄ ²	0,89	0,16	-	-	-1,61	0,11
x ₅ ²	-0,71	0,16	-	-	0,59	0,12
x ₆ ²	-0,53	0,16	-	-	-	-
x ₇ ²	-2,58	0,17	-4,91	0,98	1,95	0,11
x ₁ x ₂	-	-	-	-	-0,33	0,09
x ₂ x ₃	-	-	3,45	1,00	-0,41	0,13
x ₃ x ₆	-	-	-6,60	0,91	-	-
x ₄ x ₇	-	-	-	-	-0,45	0,17
x ₅ x ₆	0,35	0,18	-	-	-	-
Modelo	F _d (11, 69) = 130 (p _F = 0,000) R ² = 0,954		F _d (7, 73) = 38,9 (p _F = 0,000) R ² = 0,789		F _d (13, 67) = 300 (p _F = 0,000) R ² = 0,983	
Falta de ajuste	F _d (6, 63) = 1,446 (p _F = 0,211)		F _d (10, 63) = 1,493 (p _F = 0,163)		F _d (4, 63) = 1,776 (p _F = 0,145)	

Os testes F da tabela 5.17 indicam que as equações de regressão ajustadas por MQG são significativas. Quando comparam-se as equações ajustadas por MQG com as produzidas por MQO (tabela 5.15), observam-se substanciais aumentos nos coeficientes R² e reduções nos erros padrão, o que sugere a conveniência da aplicação do MQG neste caso. Vale ressaltar, contudo, que estas estatísticas não levam em conta os erros de predição das variâncias. Logo, os resultados apresentados na tabela 5.17 podem ser mais otimistas do que os valores reais.

A figura 5.15 apresenta o gráfico dos resíduos do ajuste por MQG sobre os dados da retração linear.

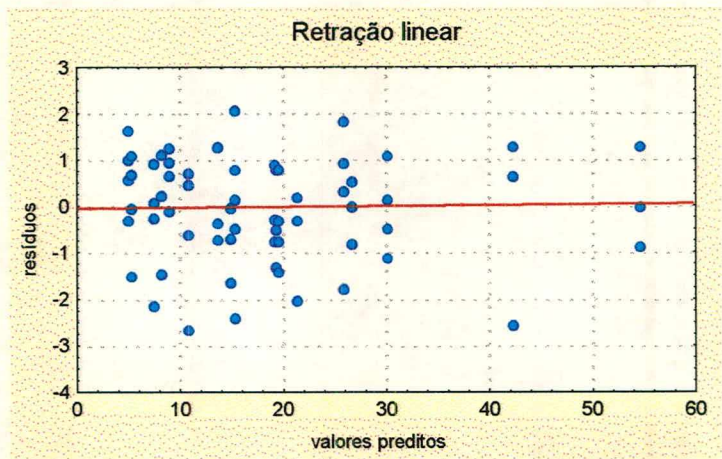


Fig. 5.15 - Análise dos resíduos do modelo da média da retração linear, ajustado por MQG.

Observa-se, na figura 5.15, uma distribuição de resíduos mais bem comportada, quando comparada com a distribuição dos resíduos resultante da aplicação de MQO (figura 5.13). Em especial, houve um acomodamento do valor discrepante observado num dos resíduos de MQO.

5.4.4 - Estimação dos parâmetros dos modelos de forma sistêmica

Para verificar se existe vantagens em adotar a abordagem de sistemas de equações de regressão aparentemente não relacionadas (SUR), calcularam-se os coeficientes de correlação de Pearson entre os resíduos dos modelos das médias e entre os resíduos dos modelos das variâncias, ambos em termos das variáveis transformadas. A tabela 5.18 apresenta estes coeficientes.

Tabela 5.18 - Coeficientes de correlação entre os resíduos dos modelos das médias e entre os resíduos dos modelos das variâncias.

equações	modelos das médias			modelos das variâncias		
	(1)	(2)	(3)	(1)	(2)	(3)
(1)		0,43	-0,70		0,12	0,11
(2)	0,43		-0,46	0,12		-0,04
(3)	-0,70	-0,46		0,11	-0,04	

Conforme se observa na tabela 5.18, as correlações amostrais entre os resíduos das equações das médias mostraram-se razoavelmente grandes. Além disso, todas são significativamente diferentes de zero, ao nível de significância de 5%. Considerando, também, que as equações são formadas por vários termos diferentes, como mostram as tabelas 5.15 e 5.17, a aplicação do método MQG-SUR, conforme

discutido no capítulo 4, deve fornecer estimativas mais precisas do que as obtidas por análises univariadas. Por outro lado, as correlações entre os termos de erro das equações das variâncias são bastante baixas e, portanto, não se deve aplicar a abordagem sistêmica para estas equações.

A tabela 5.19 apresenta as estimativas dos parâmetros dos modelos das médias, aplicando-se o algoritmo desenvolvido no capítulo 4. O programa computacional é apresentado no apêndice.

Tabela 5.19 - Sistema de equações de regressão das médias, com estimativas obtidas pelo método MQG-SUR. Resultados após duas iterações.

Efeito	Retração linear		Resistência mecânica		Absorção de água	
	coef	erro padrão	coef	erro padrão	coef	erro padrão
intercepto	11,73	0,16	45,49	0,67	3,12	0,14
x_2	-0,72	0,07	-6,01	0,56	0,91	0,06
x_3	-0,90	0,06	-5,01	0,51	1,38	0,04
x_5	-0,83	0,06	-5,19	0,47	1,00	0,05
x_6	0,40	0,06	4,47	0,42	-1,67	0,05
x_7	-	-	-	-	0,40	0,05
x_2^2	-1,14	0,12	-	-	1,70	0,08
x_3^2	1,09	0,09	-	-	-0,91	0,09
x_4^2	0,68	0,10	-	-	-1,50	0,07
x_5^2	-0,84	0,10	-	-	0,52	0,08
x_6^2	-0,40	0,10	-	-	-	-
x_7^2	-2,75	0,11	-5,05	0,82	2,08	0,07
x_1x_2	-	-	-	-	-0,29	0,05
x_2x_3	-	-	3,13	0,85	-0,31	0,08
x_3x_6	-	-	-7,14	0,77	-	-
x_4x_7	-	-	-	-	-0,52	0,10
x_5x_6	0,10	0,11	-	-	-	-

Comparando os erros padrão das estimativas por MQG-SUR (tabela 5.19) com os erros padrão de estimativas univariadas (tabela 5.17), verifica-se que o método sistêmico produziu estimativas mais precisas. Observa-se que, no cálculo dos erros padrão, apresentados na tabela 5.19, não se leva em conta as variações das predições das variâncias e das estimativas das correlações. Contudo, considerando que no método sistêmico os valores dos erros padrão foram bem menores do que nos métodos univariados, tem-se uma provável superioridade na abordagem sistêmica. Além disso, a presente situação assemelha-se com as situações em que os estudos de simulação de Monte Carlo, realizados no capítulo 4, apontavam para a supremacia do método MQG-SUR.

As correlações calculadas após o ajuste dos modelos indicados na tabela 5.19 foram iguais a 0,53, -0,80 e -0,50, relativas aos termos de erro dos pares (y_1, y_2) , (y_1, y_3) e (y_2, y_3) , respectivamente. Ou seja, após o processo iterativo, as estimativas das correlações aumentaram de magnitude, o que reforça a conveniência do método sistêmico para o presente caso.

CONSIDERAÇÕES FINAIS

Este capítulo descreve as principais conclusões, destacando a relevância do tema e os itens de originalidade do trabalho. Em seguida, arrolam-se alguns tópicos que podem ser explorados em futuras pesquisas.

6.1 - CONCLUSÕES

A utilização das técnicas de planejamento de experimentos têm sido aplicadas em escala crescente nos programas de melhoria da qualidade, como forma de aprimorar a qualidade sem aumentar os custos de produção. A partir da década de 80, houve um aumento considerável da aplicação destas técnicas na área de qualidade, devido à introdução dos conceitos de engenharia relativos a projeto de parâmetros ou projeto robusto, divulgados pelo engenheiro japonês G. Taguchi. Embora os chamados métodos de Taguchi orientam para o uso de algumas técnicas específicas de experimentação e de análise de dados, as técnicas clássicas de planejamento e análise de experimentos podem incorporar os conceitos de Taguchi de forma consistente e mais eficiente, conforme atestam Box (1988), Myers e Montgomery (1995) e muitos outros. O presente trabalho segue esta abordagem.

Nesta linha, torna-se necessário, a partir dos dados oriundos de experimentos estatisticamente planejados, identificar fatores que alteram a média e fatores que alteram a variância do processo. Além disso, para realizar a otimização do produto ou do processo, torna-se relevante a construção de modelos matemáticos que relacionam os fatores do processo com a média e com a variância das principais características funcionais do produto. Enquanto relações em termos da média podem ser feitas pelo método usual de regressão por mínimos quadrados, os modelos em termos da variância não têm uma teoria consagrada.

Na construção de modelos para a variância, observa-se que grande parte dos trabalhos aplicados baseiam-se no artigo clássico de Bartlett e Kendall (1946), o qual sugere calcular a variância amostral em cada ponto experimental ensaiado com replicações. Após uma transformação logarítmica, usam-se estas estatísticas como se fossem observações numa análise de regressão. Tal procedimento é eficiente somente quando o número de replicações for grande (dez ou mais, segundo os próprios autores).

Quando o número de replicações é pequeno, como é comum na prática, o presente trabalho sugere três procedimentos:

- a) usar fatores não significativos como se fossem replicações (*pseudo-replicações*);
- b) combinar variâncias amostrais com resíduos quadráticos e
- c) efetuar uma alteração na transformação logarítmica.

O primeiro procedimento baseia-se no trabalho de Box e Meyer (1986a), que apresenta um método de identificação de fatores de dispersão em experimentos fatoriais do tipo 2^k não replicados. Segundo este método, ao analisar um fator, consideram-se todos os outros como *pseudo-replicações*. Nair e Pregibon (1988) ampliam o método de Box e Meyer para experimentos 2^k replicados e mostram que este método só é adequado quando se tem apenas um fator influenciando fortemente a variância. No presente trabalho, este método é usado na fase de estimação e considera-se como *pseudo-replicações* somente os fatores identificados como não significativos numa fase anterior.

Na combinação de variâncias amostrais com resíduos quadráticos, usam-se estatísticas de adequação do modelo da média e conceitos da teoria de conjuntos difusos, como forma de se agregar dois métodos descritos na literatura: (1) modelo da variância construído independentemente do modelo da média e (2) estimação conjunta dos parâmetros dos modelos da média e da variância. Enquanto que no primeiro caso não se corre o risco de se ter um modelo para a variância ruim devido a um possível ajuste inadequado do modelo da média, no segundo caso tem-se um procedimento mais eficiente para a estimação dos parâmetros. O procedimento proposto neste trabalho procura reunir as vantagens dos dois métodos.

Para se estabelecer a equação de regressão em função das variâncias amostrais, dos resíduos quadráticos ou das medidas que agregam ambos, torna-se necessário uma transformação para que estas estatísticas satisfaçam aproximadamente as suposições de um modelo de regressão. A transformação logarítmica normalmente é usada para este fim. Este trabalho propõe somar uma certa quantidade antes de se proceder a transformação, evitando-se, assim, o surgimento de valores extremamente pequenos devido as características da função logarítmica. Enquanto que nos trabalhos práticos isto as vezes é feito de forma empírica, o presente trabalho cria instrumentos mais gerais para se adequar a transformação logarítmica ao modelo da variância.

Os três procedimentos aqui sugeridos são avaliados através de simulações do tipo Monte Carlo, realizadas sob a metodologia de planejamentos de experimentos. E, como consequência destes estudos, foi possível estabelecer as principais condições para o bom desempenho de cada procedimento e montar algumas estratégias na construção de modelos para a média e para a variância do processo.

Considerando que em muitas situações a qualidade precisa ser avaliada por mais de uma característica funcional, este trabalho também incorpora o estudo da variância do processo nos sistemas de equações de regressão do tipo SUR, incluindo os modelos das variâncias no processo multivariado de estimação. Descreve-se um algoritmo de estimação sistêmica em termos da filosofia da técnica dos mínimos quadrados generalizados, avaliando-o através de simulações do tipo Monte Carlo.

As estratégias para a construção de modelos para a variância e o algoritmo de estimação multivariada são aplicados em três estudos de caso extraídos da literatura. Os novos procedimentos mostraram-se plenamente factíveis nestas aplicações.

Outra característica a ressaltar foi a forma de utilização das simulações do tipo Monte Carlo. Este tipo de simulação tem sido largamente usada na comparação de métodos estatísticos. Neste trabalho, contudo, a sua utilização, realizada sob a metodologia de planejamento de experimentos, não só permitiu comparar métodos, mas também auxiliar na construção destes métodos. Este aspecto ficou evidente, por exemplo, na adequação da transformação logarítmica para o modelo da variância.

6.2 - SUGESTÕES PARA NOVAS PESQUISAS

Para que os métodos aqui aprimorados possam ser largamente aplicados na prática, faz-se necessário a existência de sistemas computacionais que facilitem as suas aplicações. Seria, portanto, de grande utilidade a construção de programas computacionais que incorporem os métodos descritos neste trabalho, juntamente com as técnicas usuais de análise de regressão. Estes programas podem ser inseridos em sistemas computacionais *amigáveis*, ou até mesmo em sistemas especialistas, para que possam ser usados pelos engenheiros da qualidade e do processo, os quais, em geral, não têm grande familiaridade com os aspectos teóricos dos métodos estatísticos.

Em termos mais específicos, pode-se explorar os procedimentos aqui descritos, tendo como objetivo principal a identificação de efeitos de dispersão. Neste trabalho, embora se aplicou alguns dos procedimentos propostos na fase de identificação dos efeitos, a ênfase central foi dada para o problema da estimação dos parâmetros.

Outra questão que pode ser investigada é a possibilidade de se conseguir uma forma mais simples para se adequar a transformação sobre as variâncias amostrais ou sobre os resíduos quadráticos. Embora se obteve uma forma objetiva de se corrigir a transformação logarítmica, ela é muito detalhista e depende demasiadamente de parâmetros desconhecidos, o que dificulta a sua aplicação prática.

Todos os procedimentos sugeridos neste trabalho foram testados em termos de simulações do tipo Monte Carlo, com a parte aleatória gerada segundo a distribuição normal. Não se estudou o desempenho destes procedimentos sob outras distribuições de probabilidade e, também, na presença de valores discrepantes. Um estudo desta natureza permite avaliar a robustez das técnicas, característica bastante desejável nos métodos estatísticos.

Este trabalho utiliza-se dos métodos de mínimos quadrados ordinários e mínimos quadrados generalizados para estimar os parâmetros dos modelos de regressão. Muitos trabalhos fazem uso da abordagem da máxima verossimilhança, ou da máxima verossimilhança restrita, para a estimação conjunta dos parâmetros dos modelos da média e da variância. Neste contexto, seria interessante uma análise de

eficiência e de robustez entre as técnicas descritas neste trabalho e o método da máxima verossimilhança ou da máxima verossimilhança restrita.

O trabalho enfatizou a importância de identificar fatores que alteram a média e fatores que alteram a variância do processo, construindo modelos para estas duas características da resposta do sistema. De certa forma, esta ênfase está associada ao modelo normal, onde a média e a variância são parâmetros independentes e caracterizam totalmente a distribuição dos dados. É a linha de pesquisa apontada por Box et al. (1978) e Box (1988), que sugere transformar os dados para aproximá-los da distribuição normal ou para separar melhor os efeitos de locação e de dispersão. Muitas vezes, as principais características de qualidade estão associadas a certas distribuições de probabilidade, onde existe relação entre a média e a variância. Nestes casos, pode-se construir modelos para apenas uma das características da distribuição, conforme abordam Hamada e Nelder (1997). Esta discussão pode ser aprofundada em futuras pesquisas.

Outra questão que pode ser mais explorada em futuras investigações é o problema de múltiplas respostas. Na análise do terceiro estudo de caso, procurou-se indentificar os efeitos significativos, analisando-se cada resposta individualmente. Somente no processo de estimação é que se considerou a abordagem multivariada. Alternativamente, poder-se-ia pesquisar a significância dos fatores sobre o sistema de respostas como um todo, resultando num conjunto de equações com as mesmas variáveis independentes. Uma análise comparativa destas duas abordagens, sob diferentes projetos experimentais, pode ser objeto de novos estudos.

As inovações desta pesquisa concentraram-se na construção de modelos. Porém, nos estudos de múltiplas respostas, o algoritmo proposto para a estimação de parâmetros dos modelos das médias e das variâncias também estima a matriz de correlações entre os termos de erro das respostas. Esta matriz de correlações pode ser usada, na fase de otimização, para especificar funções perda multivariadas que consideram as distâncias entre o vetor das características funcionais e o vetor de valores do projeto, também como função das correlações entre as características funcionais. Mas as vantagens e desvantagens desta abordagem devem ser melhores investigadas.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMES, A. E., MATTUCCI, N., MACDONALD, S., SZONYI, G., HAWKINS, D. M. - *Quality loss functions for optimization across multiple response surfaces*. Journal of Quality Technology, v. 29, n. 3, 1997, p. 339 - 346.
- AITKIN, M. - *Modelling variance heterogeneity in normal regression using GLIM*. Applied Statistics, v. 36, n. 3, 1987, p. 332 - 339.
- AMEMIYA, T. - *A note on a heteroscedastic model*. Journal of Econometrics, v. 6, 1977, p. 365 - 370.
- ATKINSON, A. C. - *Diagnostic tests for transformations*. Technometrics, v. 28, n. 1, 1986, p. 29 - 37.
- ATKINSON, A. C., COOK, R. D. - *D - optimum designs for heteroscedastic linear models*. Journal of the American Statistical Association, v. 90, n. 429, 1995, p. 204 - 212.
- BARBETTA, P. A. - *Estatística aplicada às ciências sociais*. 2 ed. Florianópolis: Editora da UFSC, 1998.
- BARTLETT, M. S., KENDALL, D. J. - *The statistical analysis of variance-heterogeneity and the logarithmic transformation*. Journal of the Royal Statistical Society, Série B, v. 8, 1946, p. 128-138.
- BAZARAA, M. S. - *Nonlinear programming: theory and algorithms*. 2 ed. USA: John Wiley & Sons, 1993.
- BERGMAN, B., HYNÉN, A. - *Dispersion effects from unreplicated designs in the 2^{k-p} series*. Technometrics, v. 39, n. 2, 1997, p. 191-198.
- BOX, G. E. P. - *Signal-to-noise ratios, performance criteria, and transformation* (com discussão). Technometrics, v. 30, n. 1, 1988, p. 1 - 40.
- BOX, G. E. P., HUNTER, W. G., HUNTER, J. S. - *Statistics for experimenters*. USA: John Wiley & Sons, 1978.
- BOX, G. E. P., HUNTER, W. G., MACGREGOR, J. F., ERJAVEC, J. - *Some problems associated with the analysis of multiresponse data*. Technometrics, v. 15, n. 1, 1973, p. 33-51.
- BOX, G. E. P., MEYER, R. D. - *Dispersion effect from fractional designs*. Technometrics, v. 28, n. 1, 1986a, p. 19-27.
- BOX, G. E. P., MEYER, R. D. - *An analysis for unreplicated fractional factorials*. Technometrics, v. 28, n. 1, 1986b, p. 11 - 18.

- CARROLL, R. J., RUPPERT, D. - *A comparison between maximum likelihood and generalized least squares in a heteroscedastic linear models*. Journal of the American Statistical Association, v. 77, n. 380, 1982a, p. 878 - 882.
- CARROLL, R. J., RUPPERT, D. - *Robust estimation in heteroscedastic linear models*. The Annals of Statistics, v. 10, n. 2, 1982b, p. 429-441.
- CARROLL, R. J., RUPPERT, D. - *Transformation in regression: a robust analysis*. Technometrics, v. 27, n. 1, 1985, p. 1 - 12.
- CARROLL, R. J., RUPPERT, D. - *Transformation and weighting in regression*. USA: Chapman and Hall, 1988.
- CARROLL, R. J., CLINE, D. B. H. - *An asymptotic theory for weighted least-squares with weights estimated by replication*. Biometrika, v. 75, n. 1, 1988, p. 35 - 43.
- CARROLL, R. J., WU, C. F. J., RUPPERT, D. - *The effect of estimating weights in weighted least squares*. Journal of the American Statistical Association, v. 83, n. 404, 1988, p. 1045 - 1054.
- CASTILHO, E. D. - *Multiresponse Process optimization via constrained confidence regions*. Journal of Quality Technology, v. 28, n. 1, 1996, p. 61 - 70.
- CASTILHO, E. D., FAN, S. K., SEMPLE, J. - *The computation of global optima in dual response systems*. Journal of Quality Technology, v. 29, n. 3, 1997, p. 347 - 353.
- CASTILHO, E. D., MONTGOMERY, D. C. - *A nonlinear programming solution to the dual response problem*. Journal of Quality Technology, v. 25, n. 3, 1993, p. 199 - 204.
- CASTILHO, E. D., MONTGOMERY, D. C., McCARVILLE, D. R. - *Modified desirability functions for multiple response optimization*. Journal of Quality Technology, v. 28, n. 3, 1996, p. 337 - 345.
- CATEN, C. S. - *Método de otimização de produtos e processos medidos por múltiplas características de qualidade*. Dissertação (Mestrado em Eng. de Produção) - PPGEP / UFRGS, Porto Alegre, 1995.
- CATEN, C. S., RIBEIRO, J. L. D. - *Etapas na otimização experimental de produtos e processos: discussão e estudo de caso*. Produção, v. 6, n. 1, 1996, p. 45 - 64.
- CHAN, L. K., MAK, T. K. - *A regression approach for discovering small variation around a target*. Applied Statistics, v. 44, n. 3, 1995, p. 369 - 377.

- CHAGANTY, N. R. - *An alternative approach to the analysis of longitudinal data via generalized estimating equations*. Journal of Statistics Planning and Inference. v. 63, 1996, p. 39 - 54.
- CHAO, P. Y., HWANG, Y. D. - *An improved Taguchi's methods in design of experiments for milling CFRP composite*. International Journal of Production Research, v. 35, n. 1, 1997, p. 51-56.
- CHATTERJEE, S., PRICE, B. - *Regression analysis by examples*. USA: John Wiley & Sons, 1977.
- CHEN, J., SHAO, J. - *Iterative weighted least squares estimators*. The Annals of Statistics, v. 21, n. 2, 1993, p. 1071-1092.
- COCHRAN, W. G., COX, G. M. - *Experimental Design*. 2 ed. USA: John Wiley & Sons, 1957.
- COLEMAN, D. E., MONTGOMERY, D. C. - *A systematic approach to planning for a design industrial experiment* (com discussão). Technometrics, v. 35, n. 1, 1993, p. 1 - 27.
- COOK, R. D. - *On the interpretation of regression plots*. Journal of the American Statistical Association, v. 89, n. 425, 1994.
- COOK, R. D., WEISBERG, S. - *Regression diagnostics with dynamic graphics* (com discussão). Technometrics, v. 31, n. 3, 1989, p. 277 - 311.
- COPPELAND, K. A. F., NELSON, P. R. - *Dual response optimization via direct function minimization*. Journal of Quality Technology, v. 28, n. 3, 1996, p. 331 - 336.
- CORDEIRO, G. - *Introdução à teoria de verossimilhança*. 10º Simpósio Nacional de Probabilidade e Estatística, Rio de Janeiro - RJ, 1992.
- DAVIDIAN, M., CARROLL, R. J. - *Variance function estimation*. Journal of the American Statistical Association, v. 82, n. 400, 1987, p. 1079-1091.
- DERRINGER, G., SUICH, R. - *Simultaneous optimization of several response variables*. Journal of Quality Technology, v. 12, n. 4, 1980, p. 214 - 219.
- DRAPER, N. R., SMITH, H. - *Applied regression analysis*. 2 ed. USA: John Wiley & Sons, 1981.
- ELSAYED, E. A., CHEN, A. - *Optimal levels of process parameters for products with multiple characteristics*. International Journal of Production Research, v. 31, n. 5, 1993, p. 1117 - 1132.

- ENGEL, J. - *Modelling variation in industrial experiments*. Applied Statistics, v. 41, n. 3, 1992, p. 579-593.
- ENGEL, J., HUELE, A. F. - *A generalized linear modeling approach to robust design*. Technometrics, v. 38, n. 4, 1996, p. 365 - 373.
- FERRER, A. J., ROMERO, R. - *Small samples estimation of dispersion effects from unreplicated data*. Communications in Statistics. Part B: Simulation and Computation, v. 22, n. 4, 1993, p.975 - 995.
- FERRER, A. J., ROMERO, R. - *A simple method to study dispersion effects from non necessarily replicated data in industrial context*. Quality Engineering, v. 7, n. 4, 1995, p. 747 - 755.
- FIOD NETO, M. - *Taguchi e a melhoria da qualidade: uma releitura crítica*. Florianópolis: Editora da UFSC, 1997.
- FISHER, R. A. - *The design of experiments*. 6 ed. England: Oliver and Boyd, 1951.
- GILTINAN, D. M., CARROLL, R. J., RUPPERT, D. - *Some new estimation methods for weight regression when there are possible outliers*. Technometrics, v. 28, n. 3, 1986, p. 219 - 230.
- GNANADESIKAN, R. - *Methods for statistical data analysis of multivariate observations*. USA: John Wiley & Sons, 1977.
- GRECO, J. M. - *Generalized linear models and process variation*. Journal of Quality Technology, v. 25, n. 4, 1993, p. 288 - 295.
- GUEDES, T. A. -. *Procedimentos de otimização no planejamento e controle da qualidade de produtos e processos*. Tese (Doutorado em Eng. de Produção) CPGEF / UFSC, 1996.
- HAMADA, M., NELDER, J. A. - *Generalized linear models for quality-improvement experiments*. Journal of Quality Technology, v. 29, n. 3, 1997, p. 292 - 304.
- HAMADA, M., WU, C. F. J. - *Analysis of designed experiments with complex aliasing*. Journal of Quality Technology, vol. 24, n. 3, 1992, p. 130-137.
- HARVILLE, D. A. - *Maximum likelihood approaches to variance component estimation and related problems*. Journal of the American Statistical Association, v. 72, n. 358, 1977, p. 320 - 340.
- HARVEY, A. C. - *Estimating regression models with multiplicative heteroscedascity*. Econometrica, v. 44, n. 3, 1976, p. 461 - 465.

- HILL, W. J., HUNTER, W.G. - *A review of response surface methodology: a literature review*. Technometrics, v. 8, 1966, p. 571-590
- HOOVER, P. M. - *Iterative weighted least squares estimation in heterocedastic linear models*. Journal of the American Statistical Association, v. 88, n. 421, 1993, p. 179-184.
- HUGHES-OLIVER, J. M., LU, J. C., DAVIS, J. C., GYURCSIK, R. S. - *Achieving uniformity in a semiconductor fabrication process using spatial modeling*. Journal of the American Statistical Association, v. 93, n. 441, 1998, p. 36 - 45.
- JOHNSON, R. A., WICHERN, D. W. - *Applied multivariate statistical analysis*. 3 ed. USA: Prentice Hall, 1992.
- JUDGE, G. G., HILL, R. C., GRIFFITHS, W. E., LUTKEPOHL, H., LEE, T. C. - *Introduction to the theory and practice of econometrics*, 2nd edition. USA: John Wiley & Sons, 1988.
- JURAN, J. M., GRZYNA, F. M. - *Controle da qualidade: handbook*. V. 1, São Paulo: Mc. Graw-Hill, 1991.
- KACKAR, R. N. - *Taguchi's quality philosophy: analysis and commentary*. Quality Progress, December, 1986, p. 21 - 29.
- KAKWANI, N. C. - *The unbiasedness of Zellner's seemingly unrelated regressions equations estimators*. Journal of the American Statistical Association, v. 62, 1967, p. 141 - 142.
- KAPUR, C. K., CHO, B. R. - *Economic design of the specification region for multiple quality characteristics*. IIE Transactions, v. 28, 1996, p. 237 - 248.
- KHATTREE, R., NAIK, D. N. - *Applied multivariate statistics with SAS software*. USA: SAS Institute, 1995.
- KHURI, A. I., CORNELL, J. A. - *Response surfaces: designs and analyses*. 2 ed. USA: Marcel Dekker, 1996.
- KIM, C., STORER, B. E., JEONG, M. - *A note on Box-Cox transformation diagnostics*. v. 38, n. 2, 1996, p. 178 - 180.
- KLIR, G. J., YUAN, B. - *Fuzzy sets and fuzzy logic*. USA: Prentice Hall, 1995.
- LAW, A. M., KELTON, W. D. - *Simulation Modeling and Analysis*. USA: Mc Graw Hill, 1982.
- LENTH, R. V. - *Quick and easy analysis of unreplicated factorials*. Technometrics, v. 31, n. 4, 1989, p. 469 - 473.

- LÉON, R. V., SHOEMAKER, A. C., KACKER, R. N. - *Performance measures independent of adjustment* (com discussão). *Technometrics*, v. 29, n. 3, 1987, p. 253 - 285.
- LIN, D. K. J., TU, W. - *Dual response surface optimization*. *Journal of Quality Technology*, v.27, n. 1, 1995.
- LIU, X., DOOLEY, K. J., ANDERSON, J. C. - *Combining process knowledge for continuous quality improvement*. *IIE Transaction*, n. 27, 1995, p. 811-819.
- LUCAS, J. M. - *How to achieve a robust process using response surface methodology*. *Journal of Quality Technology*, v. 26, n. 4, 1994, p. 248 - 260.
- LUNER, J. J. - *Achieving continuous improvement with the dual response approach: a demonstration of the roman catapult*. *Quality Engineering*, v. 6, n. 4, 1994, p. 691 - 705.
- MAGNUS, J. R. - *Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix*. *Journal of Econometrics*, v. 7, 1978, p. 281 - 312.
- MANDY, D. M., MARTINS-FILHO, C. - *Seemingly unrelated regressions under additive heterocedasticity*. *Journal of Econometrics*, v. 58, 1993, p. 315-346.
- MAYS, D. P., EATER, S. M. - *Optimal response surface designs in the presence of dispersion effects*. *Journal of Quality Technology*, v. 29, n. 1, 1997, p. 59 - 70.
- MEHTA, J. S., SWAMY, V. B. - *Further evidence on the relative efficiencies of Zellner's seemingly unrelated regressions estimators*. *Journal of the American Statistical Association*, v. 71, n. 355, 1976, p. 141 - 142.
- MESENBRINK, P., LU, J. C., MCKENZIE, R., TAHERI, J. - *Characterization and optimization of wave-soldering process*. *Journal of the American Statistical Association*, v. 89, n. 428, 1994, p. 1209 - 1217.
- MONTGOMERY, D. C. - *Design and analysis of experiments*, 4 ed., USA: John Wiley & Sons, 1997.
- MYERS, R. H., CARTER, W. H. - *Response surface techniques for dual response systems*. *Technometrics*, v. 15, n. 2, 1973, p. 301 - 317.
- MYERS, R. H., MONTGOMERY, D. C. - *Response surface methodology: process and product optimization using designed experiments*. USA: John Wiley & Sons, 1995.
- MYERS, R. H. , KHURI, A. I., CARTER, W. H. - *Response surface methodology: 1966 - 1988*. *Technometrics*, v. 31, n. 2, 1989, p. 137 - 157.

- NAIR, V. N. - *Taguchi's parameter design: a panel discussion*. Technometrics, v. 34, n. 2, 1992, p. 127-161.
- NAIR, V. N., PREGIBON, D. - *A data analysis strategy for quality engineering experiments*. AT&T Technical Journal, v. 65, 1986, p. 73-84.
- NAIR, V. N., PREGIBON, D. - *Analysing dispersion effects from replicated factorial experiments*. Technometrics, v. 30, n. 3, 1988, p. 247 - 257
- NELDER, J. A., LEE, Y. - *Letters to the editor: joint modeling of mean and dispersion*. Technometrics, v. 40, n. 2, 1998, p. 168 - 175.
- OH, H. L. - *A changing paradigm in quality*. IEEE Transactions on Reliability, v. 44, n. 2, 1995, p. 265 - 270.
- OLGUÍN, J., FEARN, T. - *A new look at half-normal plots for assessing the significance of contrasts for unreplicated factorials*. Applied Statistics, v. 46, n. 4, 1997, p. 449 - 462.
- PALADINI, E. P. - *Gestão da qualidade no processo*. São Paulo: Atlas, 1995.
- PASA, G. S., RIBEIRO, J. L. D., NUNES, D. B. - *Otimização da formulação de uma mistura - parte 1: qualidade*. Produção, v. 6, n. 1, 1996, p. 65 - 82.
- PHADKE, M. S. - *Quality Engineering using robust design*. USA: Prentice Hall, 1989.
- PRESS, S. J. - *Applied multivariate analysis*. USA: Holt, Rinehart and Winston, 1972.
- RAJ, B. - *Linear regression with random coefficients: the finite sample and convergence properties*. Journal of the American Statistical Association, v. 70, 1975, p. 127-137.
- RIBEIRO, J. L. D., ELSAYED, E. A. - *A case study on process optimization using the gradient loss function*. International Journal of Production Research, v. 33, n. 12, 1995, p. 3233-3248.
- RIBEIRO, J. L. D., PASA, G. S., NUNES, D. B. - *Otimização da formulação de uma mistura - parte 2: custo*. Produção, v. 7, n. 2, 1997, p. 189 - 201.
- RIBOLDI, J., NASCIMENTO, L. C. S. C. - *Metodologia de superfície de resposta: uma abordagem introdutória*. Porto Alegre, 1994.
- ROSENBAUM, P. R. - *Some useful compound dispersion experiments in quality design*. Technometrics, v. 38, n. 4, 1996, p. 354 - 364.

- ROSS, P. J. - Aplicações das técnicas de Taguchi na engenharia de qualidade. São Paulo: Mc. Graw Hill, 1991.
- SARIN, S. - *Teaching Taguchi's approach to parameter design*. Quality Progress, May, 1997, p. 102 - 106.
- SMYTH, G. K. - *Generalized linear models with varying dispersion*. Journal of the Royal Statistics Society, v. 51, n. 1, 1989, p. 47 - 60.
- SRIVASTAVA, K. S., GILES, D. E. A. - *Seemingly unrelated regression equations models* - USA: Marcel Dekker, 1987.
- SRIVASTAVA, K. S., MAEKAWA, K. - *Efficiency properties of feasible generalized least squares estimators in SURE models under non-normal disturbance*. Journal of Econometrics, v. 66, 1995, p. 99 - 121.
- SHOEMAKER, A. C., TSUI, K. L., WU, C. F. J. - *Economical experimentation methods for robust design*. Technometrics, v. 33, n. 4, 1991, p. 415 - 427.
- STEINER, S. H., HAMADA, M. - *Making mixtures robust to noise and mixing measurement errors*. Journal of Quality Technology, v. 39, n.4, 1997, p. 441 - 450.
- STIGLER, S. M. - *The history of statistics: the measurement of uncertainty before 1900*. USA: Harward, 1986.
- TAGUCHI, G. - *System of experimental design: engineering methods to optimize quality and minimize costs*. 2 ed. USA: UNIPUB, 1987.
- TAGUCHI, G., ELSAYED, E. A., HSIANG, T. - *Engenharia da qualidade em sistemas de produção*. São Paulo: McGraw-Hill, 1990.
- TSUI, K. L. - *An overview of Taguchi method and newly developed statistical methods for robust design*. IIE Transactions, v. 24, n. 5, 1992, p. 44 - 57.
- VACCARO, G. L. R.; ISOPO, M.; RIBEIRO, J. L. D., BARBETTA, P. A. - *Um algoritmo para regressão múltipla multivariada para a modelagem de características de qualidade correlacionadas*. ENEGEP, 1996.
- VERBYLA, A. P. - *Modelling variance heterogeneity: residual maximum likelihood and diagnostics*. Journal of the Royal Statistics Society, v. 55, n. 2, 1993, p. 493 - 508.
- VINING, G. G., BOHN, L. L. - *Response surfaces for the mean and variance using a nonparametric approach*. Journal of Quality Technology, v. 30, n. 3, 1998, p. 282 - 291.
- VINING, G. G., MYERS, R. H. - *Combining Taguchi and response surface philosophies: a dual response approach*. Journal of Quality Technology, v. 22, n. 1, 1990, p. 38 - 45.

- WELCH, W. J., BUCK R. J., SACKS, J., WYNN, H. P., MITCHELL, T. J., MORRIS, M. D. - *Screening, predicting, and computer experiments*. Technometrics, v. 34, n. 1, 1992, p. 15 - 25.
- WELCH, W. J., YU, T. K., KANG, S. M., SACKS, J. - *Computer experiments for quality control by parameter design*. Journal of Quality Technology, v. 22, n. 1, 1990, p. 15 - 22.
- WHITE, H. - *A heteroscedasticity-consistent covariance matrix estimator and direct test for heteroscedasticity*. Econometrica, v. 48, n.4, 1980, p. 817 - 838.
- WOLFINGER, R. D., TOBIAS, R. D. - *Joint estimation of location, dispersion, and random effects in robust design*. Technometrics, v. 40, n. 1, 1998, p. 62 - 71.
- ZEGER, S. L., LIANG, K. Y. - *Longitudinal data analysis using generalized linear models*. Biometrika, v. 73, n. 1, 1986, p. 13 - 22.
- ZELLNER, A. - *An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias*. Journal of the American Statistical Association, v. 57, n. 298, 1962, p. 348-368.

APÊNDICE: programas computacionais

Os programas estão escritos no procedimento IML do sistema SAS, que permite programação em linguagem matricial.

CONTEÚDO

- A - Agregação de resíduos quadráticos: simulação
- B - Agregação de variâncias amostrais: simulação
- C - Distribuição $\log(\chi_1 + q)$
- D - Avaliação da transformação logarítmica modificada: simulação
- E - Agregação de variâncias amostrais e resíduos quadráticos: simulação
- F - Correlações entre logaritmos de variâncias amostrais
- G - MQG - SUR: simulação
- H - Estudo de caso nº 1
- I - Estudo de caso nº 2
- J - Estudo de caso nº 3

A - Agregação de resíduos quadráticos: simulação

Programa computacional do estudo de simulação número 1 (seção 3.2.1).
Inclui os ensaios de 1 a 32. Para os ensaios de 33 a 64 alterou-se “data xdad” para contemplar a matriz de planejamento de um experimento fatorial 2^6 .

```

** AGREGAÇÃO - MÉTODO RQ 2^4 1 REPLIC. ;
** Estudo de simulação n. 1 - parte 1;
data planej;                                /*matriz de planejamento do modelo de simulação*/
input run a b c d e f g h;                 /*nome dos fatores de simulação*/
cards;
1      1      1      1      1      1      1      1      1
2      1      1      1      1      1      -1     1      -1
3      1      1      1      1      -1     1      1      -1
4      1      1      1      1      -1     -1     1      1
5      1      1      1      -1     1      1      -1     1
6      1      1      1      -1     1      -1     -1     -1
7      1      1      1      -1     -1     1      -1     -1
8      1      1      1      -1     -1     -1     -1     1
9      1      1      -1     1      1      1      -1     1
10     1      1      -1     1      1      -1     -1     -1
11     1      1      -1     1      -1     1      -1     -1
12     1      1      -1     1      -1     -1     -1     1
13     1      1      -1     -1     1      1      1      1
14     1      1      -1     -1     1      -1     1      -1
15     1      1      -1     -1     -1     1      1      -1
16     1      1      -1     -1     -1     -1     1      1
17     1      -1     1      1      1      1      -1     -1
18     1      -1     1      1      1      -1     -1     1
19     1      -1     1      1      -1     1      -1     1
20     1      -1     1      1      -1     -1     -1     -1
21     1      -1     1      -1     1      1      1      -1
22     1      -1     1      -1     1      -1     1      1
23     1      -1     1      -1     -1     1      1      1
24     1      -1     1      -1     -1     -1     1      -1
25     1      -1     -1     1      1      1      1      -1
26     1      -1     -1     1      1      -1     1      1
27     1      -1     -1     1      -1     1      1      1
28     1      -1     -1     1      -1     -1     1      -1
29     1      -1     -1     -1     1      1      -1     -1
30     1      -1     -1     -1     1      -1     -1     1
31     1      -1     -1     -1     -1     1      -1     1
32     1      -1     -1     -1     -1     -1     -1     -1
;
data xdad;                                  /*Matriz de planejamento do expto. simulado - fatorial 2^4*/
input cexp x1-x4;
cards;
1      -1     -1     -1     -1
2      -1     -1     -1     1
3      -1     -1     1     -1
4      -1     -1     1     1
5      -1     1     -1     -1
6      -1     1     -1     1
7      -1     1     1     -1
8      -1     1     1     1
9      1     -1     -1     -1
10     1     -1     -1     1
11     1     -1     1     -1
12     1     -1     1     1
13     1     1     -1     -1

```

```

14      1      1      -1     1
15      1      1      1      -1
16      1      1      1      1
;
proc iml;                                /*procedimento SAS de linguagem matricial*/

/* PROGRAMA DE GERAÇÃO DOS DADOS */
start GERA;
  nor = j(n_amostr,1,.);                /*construção de um vetor N(0, 1) de dimensão n_amostr*/
  normal=0;
  do i=1 to n_amostr;
    call rannor(seed, normal);
    nor[i] = normal;
  end;
  erro = sqrt(vreal) # nor;              /*vetor do termo de erros*/
  resposta = yreal + erro;               /*vetor das respostas*/
finish GERA;

/* PROGRAMA AUXILIAR PARA CONSTRUÇÃO DA MATRIZ DE PLANEJAMENTO DO MODELO
DE VARIÂNCIA */
start PLAN;
  zz2 = j(ncexp,qq,.);                  /*Ajuste da matriz de planejamento zz em termos...*/
  cexp2 = j(n_amostr,1,.);              /*... do número de pontos experimentais distintos, zz2*/
  cexp2[1]=1;                           /* cexp2: condições experimentais (≠ cexp qdo fator agregado)*/
  do j=1 to qq;
    zz2[1,j] = zz[1,j];
  end;
  ii = 1;
  do i=2 to n_amostr;
    i1=i-1;
    do k=1 to nfat;
      if fat[i,k] ^= fat[i1,k] then do;
        ii=ii+1;
        do j = 1 to qq;
          zz2[ii,j] = zz[i,j];
        end;
        cexp2[i]=ii;
        k=nfat;
      end;
      else cexp2[i]=ii;
    end;
  end;
finish PLAN;

/* PROGRAMA AUXILIAR PARA REGRESSAO USANDO MQO */
start REG;
  nobs=nrow(x);                         /*número de observações*/
  npar=ncol(x);                         /*número de parâmetros*/
  xxi=inv(x`*x);
  coef=xxi*x`*y;                        /*estimativador MQO dos coeficientes*/
  yhat=x*coef;                          /*valores preditos*/
  resid = y-yhat;                       /*resíduos brutos*/
  h = j(nobs,1,.);
  do i=1 to nobs;
    h[i] = x[i,]*xxi*x[i,];              /*vetor de correção dos resíduos, devido a perda de g.l.*/
  end;
Finish REG;

/* PROGRAMA PARA ESTIMAR OS MODELOS DA MÉDIA E DA VARIÂNCIA */
start ESTIMA;
/* CÁLCULO DO NÚMERO DE REPLICAÇÕES */
  nrep=j(ncexp,1,1);                   /*vetor que vai conter o núm. de obs. em cada...*/
  if ncexp < n_amostr then do;          /*... ponto experimental (cexp) */
    j=0;
  end;

```

```

do i=1 to ncexp;
  controle=0;
  nrep[i] = 0;
  do while(controle=0);
    j=j+1;
    nrep[i]=nrep[i]+1;
    if j=n_amost then do;
      controle=1;
      goto fim;
    end;
    if cexp2[j+1] ^= cexp2[j] then controle=1;
  fim:
  end;
end;
end;

/* MODELO INICIAL PARA AS MÉDIAS - MQO */
y = resposta;
x = xx;
RUN REG;                                     /*chama o programa auxiliar REG*/
r = resid/sqrt(1-h);                         /*correção dos resíduos*/

/* MODELO INICIAL PARA AS VARIÂNCIAS */
rq = r##2;
rqm=j(ncexp, 1,.);                          /*cálculo do resíduo quadrático médio em cada pto. experimental*/
fa=0;
do i=1 to ncexp;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rq[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;
y = log(rqm);                               /*transformação logarítmica*/
x = zz2;                                    /*matriz de planejamento criada no prog. auxiliar PLAN*/
RUN REG;                                    /*chama o programa auxiliar REG (regressão)*/
v = exp(yhat);                              /*variâncias preditas*/
/*INÍCIO DO PROCESSO ITERATIVO*/
do iter=1 to niter;                         /* preparação para o modelo das médias usando MQG */
  va=j(n_amost, 1,.);                      /*correção da dimensão do vetor de variâncias preditas ...*/
  cum = 0;                                  /* ... para ser usado no MQG - modelo das médias*/
  do i=1 to ncexp;
    if i > 1 then cum = cum + nrep[i-1];
    ii = nrep[i];
    do j=1 to ii;
      va[cum+j] = v[i];
    end;
  end;
  w = 1/sqrt(va);                          /*vetor de raiz dos pesos - MQG*/
  y = resposta # w;                        /*transformação das respostas*/
  x = xx # w;                              /*transformação das variáveis independentes e constante*/
  RUN REG;                                /*REG para processar MQG*/
  xb=xx*coef;                             /*valores preditos na métrica original*/
  rnovo= (resposta-xb)/sqrt(1-h);          /*resíduos corrigidos na métrica original*/
  rqnovo=rnovo##2;
  /* modelo para as variâncias */
  rqm=j(ncexp, 1,.);                      /*cálculo do resíduo quadrático médio em cada pto. experimental*/
  fa=0;
  do i=1 to ncexp;
    soma=0;
    do j=1 to nrep[i];

```

```

    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rqnovo[jj];
end;
rqm[i]=soma/nrep[i];
fa=fa+nrep[i];
end;
y = LOG(rqm);
x = zz2;
RUN REG;
v= exp(yhat);
end;
finish ESTIMA;

/* PROGRAMA PRINCIPAL DE CADA ENSAIO SIMULADO*/
Start PROG;
qq=ncol(zz);          /*número de parâmetros do modelo da variância*/
yreal = xgera*betagera; /*parte determinística das respostas (valor esperado)*/
vreal=g*exp(zgera*tetagera); /*função da variância*/
seed=1613218064;      /*semente inicial*/
t = j(n_simula,qq,.); /*vetores que vão conter as estimativas dos ...*/
ta=t;                 /*...parâmetros dos modelos das variâncias*/
Do simula=1 to n_simula;
  run GERA;            /*chama o programa auxiliar para gerar as respostas*/
  /* modelo com variâncias obtidas pelas replicações genuínas */
  ncexp=n_cexp;
  nfat=n_fat;
  fat=fat1;
  run PLAN;
  run ESTIMA;
  t[simula,]=coef;    /*guarda as estimativas - sem agregação*/
  /* modelo com variâncias obtidas considerando o último fator como pseudo-replicação */
  ncexp=n_cexp/2;
  nfat=n_fat-1;
  fat=fat2;
  run PLAN;
  run ESTIMA;
  ta[simula,]=coef;   /*guarda as estimativas - com agregação*/
End;
tmed = t[+,]/n_simula; /*estatísticas das estimativas com e sem a agregação*/
tamed = ta[+,]/n_simula;
sq = (t-repeat(tmed,n_simula,1))#(t-repeat(tmed,n_simula,1));
sq2 = (ta-repeat(tamed,n_simula,1))#(ta-repeat(tamed,n_simula,1));
eq = (t-repeat(teta`,n_simula,1))#(t-repeat(teta`,n_simula,1));
eq2 = (ta-repeat(teta`,n_simula,1))#(ta-repeat(teta`,n_simula,1));
eqm=eq[+,]/n_simula; /* erro quadrático médio - sem agregação*/
eqmag=eq2[+,]/n_simula; /* erro quadrático médio - com agregação*/
efic=eqm/eqmag;       /*eficiência "com" em relação "sem" agregação*/
var = sq[+,]/(n_simula-1); /*variância das estimativas*/
var2 = sq2[+,]/(n_simula-1);
vic = tmed - teta`;    /*vício das estimativas*/
vic2 = tamed - teta`;
ef[ensaio] = efic[2];  /*armazena apenas as estatísticas do parâmetro focal*/
var1[ensaio] = var[2];
var_ag[ensaio] = var2[2];
vic1[ensaio] = vic[2];
vic_ag[ensaio] = vic2[2];
finish PROG;

/* META PROGRAMA PARA EFETUAR A SIMULAÇÃO SOB UM PLANEJAMENTO EXPERIMENTAL */
start META;
n_ensaio=32;          /*definição do número de ensaios da simulação*/
ef=j(n_ensaio,1,.); var1 = ef; var_ag=ef; vic1=ef; vic_ag=ef;
use planej;

```

```

read all var{run a b c d e f g h} into planej;
n_amost = 16; /*tamanho da amostra simulada*/
n_cexp = 16; /*número de pontos experimentais em cada amostra simulada*/
n_simula = 4000; /*número de simulações de dada ensaio*/
niter=2; /*número de iterações no processo de estimação*/
n_fat=4; /*número de fatores*/
yreal=j(n_amost, 1,.);
vreal=j(n_amost, 1,.);
um = j(n_amost,1,1);
use xdad;
read all var {cexp x1 x2 x3 x4};
fat1 = x1 || x2 || x3 || x4; /*matriz de planejamento sem agregação*/
fat2 = x1 || x2 || x3; /*matriz de planejamento com agregação*/
x12=x1#x2;
x13=x1#x3;
x23=x2#x3;
/*****/
ensaio=1; /*condições relativas ao ensaio 1 (nível +1 em todos os fatores)*/
g=3; /*fator G*/
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23; /*fator B e D*/
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23; /*fator C*/
betagera={8, 8,4,8, 8, 4,4,4}; /*fator H*/
tetagera={0, .896,.347,.347, .347, .347,.347,.347}; /*fator E e F*/
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23; /*fator B e D*/
zz = um || x1 || x2 || x3 || x12 || x13 || x23; /*fator C*/
teta={0, .896,.347,.347, .347,.347,.347}; /*fator F*/
run PROG;
/*****/
ensaio=2;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 1, .5,.5,.5};
tetagera={0, .549,.347,.347, .347, .347,.347,.347};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .549,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=3;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 1, .5,.5,.5};
tetagera={0, .896,.347,.347, 0, .347,.347,.347};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .896,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=4;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 8, 4,4,4};
tetagera={0, .549,.347,.347, 0, .347,.347,.347};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .549,.347,.347, .347,.347,.347};
qq=ncol(zz);
yreal = xgera*betagera;
vreal=g*exp(zgera*tetagera);
run PROG;
/*****/

```



```

ensaio=5;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 0, 4,4,4};
tetagera={0, .896,.347,.347, .347, .347,.347,.347};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .896,.347,.347, .347,.347,.347};
run PROG;
/*****/

ensaio=6;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 0, .5,.5,.5};
tetagera={0, .549,.347,.347, 0, .347,.347,.347};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .549,.347,.347, .347,.347,.347};
run PROG;
/*****/

ensaio=7;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 0, .5,.5,.5};
tetagera={0, .896,.347,.347, 0, .347,.347,.347};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .896,.347,.347, .347,.347,.347};
run PROG;
/*****/

ensaio=8;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 0, 4,4,4};
tetagera={0, .549,.347,.347, 0, .347,.347,.347};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .549,.347,.347, .347,.347,.347};
run PROG;
/*****/

ensaio=9;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 8, 4,4,4};
tetagera={0, .896,0,0, .347, 0,0,0};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1;
teta={0, .896};
run PROG;
/*****/

ensaio=10;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 1, .5,.5,.5};
tetagera={0, .549,0,0, .347, 0,0,0};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1;
teta={0, .549};

```

```

run PROG;
/*****/
ensaio=11;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 1, .5,.5,.5};
tetagera={0, .896,0,0, 0, 0,0,0};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1;
teta={0, .896};
run PROG;
/*****/
ensaio=12;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 8, 4,4,4};
tetagera={0, .549,0,0, 0, 0,0,0};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1;
teta={0, .549};
run PROG;
/*****/
ensaio=13;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 0, 4,4,4};
tetagera={0, .896,0,0, .347, 0,0,0};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1;
teta={0, .896};
run PROG;
/*****/
ensaio=14;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 0, .5,.5,.5};
tetagera={0, .547,0,0, .347, 0,0,0};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1;
teta={0, .547};
run PROG;
/*****/
ensaio=15;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 0, .5,.5,.5};
tetagera={0, .896,0,0, 0, 0,0,0};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1;
teta={0, .896};
run PROG;
/*****/
ensaio=16;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 0, 4,4,4};
tetagera={0, .549,0,0, 0, 0,0,0};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;

```

```

zz = um || x1;
teta={0, .549};
run PROG;
/*****/
ensaio=17;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 1, 0,0,0};
tetagera={0, .896,.347,.347, .347, .347,.347,.347};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .896,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=18;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 8, 0,0,0};
tetagera={0, .549,.347,.347, .347, .347,.347,.347};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .549,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=19;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 8, 0,0,0};
tetagera={0, .896,.347,.347, 0, .347,.347,.347};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .896,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=20;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 1, 0,0,0};
tetagera={0, .547,.347,.347, 0, .347,.347,.347};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .547,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=21;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 0, 0,0,0};
tetagera={0, .896,.347,.347, .347, .347,.347,.347};
xx = um || x1 || x2 || x3;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .896,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=22;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 0, 0,0,0};

```

```

tetagera={0, .549,.347,.347, .347, .347,.347,.347};
xx = um || x1 || x2 || x3;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .549,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=23;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 0, 0,0,0};
tetagera={0, .896,.347,.347, 0, .347,.347,.347};
xx = um || x1 || x2 || x3;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .896,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=24;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 0, 0,0,0};
tetagera={0, .549,.347,.347, 0, .347,.347,.347};
xx = um || x1 || x2 || x3;
zz = um || x1 || x2 || x3 || x12 || x13 || x23;
teta={0, .549,.347,.347, .347,.347,.347};
run PROG;
/*****/
ensaio=25;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 1, 0,0,0};
tetagera={0, .896,0,0, .347, 0,0,0};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
teta={0, .896};
run PROG;
/*****/
ensaio=26;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 8, 0,0,0};
tetagera={0, .549,0,0, .347, 0,0,0};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
teta={0, .549};
run PROG;
/*****/
ensaio=27;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 8, 0,0,0};
tetagera={0, .896,0,0, 0, 0,0,0};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
teta={0, .896};
run PROG;
/*****/
ensaio=28;
g=3;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;

```

```

zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 1, 0,0,0};
tetagera={0, .549,0,0, 0, 0,0,0};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
teta={0, .549};
run PROG;
/*****/
ensaio=29;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 0, 0,0,0};
tetagera={0, .896,0,0, .347, 0,0,0};
xx = um || x1 || x2 || x3;
zz = um || x1;
teta={0, .896};
run PROG;
/*****/
ensaio=30;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 0, 0,0,0};
tetagera={0, .549,0,0, .347, 0,0,0};
xx = um || x1 || x2 || x3;
zz = um || x1;
teta={0, .549};
run PROG;
/*****/
ensaio=31;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={8, 8,4,8, 0, 0,0,0};
tetagera={0, .896,0,0, 0, 0,0,0};
xx = um || x1 || x2 || x3;
zz = um || x1;
teta={0, .896};
run PROG;
/*****/
ensaio=32;
g=1;
xgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zgera = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
betagera={0, 1,1,1, 0, 0,0,0};
tetagera={0, .549,0,0, 0, 0,0,0};
xx = um || x1 || x2 || x3;
zz = um || x1;
teta={0, .549};
run PROG;
/*****/
print planej[format=3.0] eff[format=5.3] var1[format=5.3] var_ag[format=5.3]
      vic1[format=5.3] vic_ag[format=5.3];
finish META;
RUN META;

```

B - Agregação de variâncias amostrais: simulação

Programa computacional do estudo de simulação número 2 (seção 3.2.2).

Inclui os ensaios de 33 a 64. Nos ensaios de 1 a 32, altera-se “data xdad”. Os textos

entre vários asteriscos indicam a inclusão de algum programa auxiliar de casos anteriores.

** AGREGAÇÃO - MÉTODO RQ e S2 2^4 2 REPLIC. ;

** Estudo de simulação n. 2 - parte 2;

data planej;

/* matriz de planejamento da simulação*/

input run a b c d e;

cards;

33	-1	-1	-1	-1	-1
34	-1	-1	-1	-1	1
35	-1	-1	-1	1	-1
36	-1	-1	-1	1	1
37	-1	-1	1	-1	-1
38	-1	-1	1	-1	1
39	-1	-1	1	1	-1
40	-1	-1	1	1	1
41	-1	1	-1	-1	-1
42	-1	1	-1	-1	1
43	-1	1	-1	1	-1
44	-1	1	-1	1	1
45	-1	1	1	-1	-1
46	-1	1	1	-1	1
47	-1	1	1	1	-1
48	-1	1	1	1	1
49	1	-1	-1	-1	-1
50	1	-1	-1	-1	1
51	1	-1	-1	1	-1
52	1	-1	-1	1	1
53	1	-1	1	-1	-1
54	1	-1	1	-1	1
55	1	-1	1	1	-1
56	1	-1	1	1	1
57	1	1	-1	-1	-1
58	1	1	-1	-1	1
59	1	1	-1	1	-1
60	1	1	-1	1	1
61	1	1	1	-1	-1
62	1	1	1	-1	1
63	1	1	1	1	-1
64	1	1	1	1	1

;

data xdad;

/*matriz experimental do experimento simulado*/

input cexp x1-x4;

cards;

1	-1	-1	-1	-1
2	-1	-1	-1	1
3	-1	-1	1	-1
4	-1	-1	1	1
5	-1	1	-1	-1
6	-1	1	-1	1
7	-1	1	1	-1
8	-1	1	1	1
9	1	-1	-1	-1
10	1	-1	-1	1
11	1	-1	1	-1
12	1	-1	1	1
13	1	1	-1	-1
14	1	1	-1	1
15	1	1	1	-1
16	1	1	1	1
1	-1	-1	-1	-1
2	-1	-1	-1	1
3	-1	-1	1	-1
4	-1	-1	1	1
5	-1	1	-1	-1
6	-1	1	-1	1
7	-1	1	1	-1
8	-1	1	1	1

9	1	-1	-1	-1
10	1	-1	-1	1
11	1	-1	1	-1
12	1	-1	1	1
13	1	1	-1	-1
14	1	1	-1	1
15	1	1	1	-1
16	1	1	1	1

```

proc sort; by cexp;          /*ordenação das linhas da matriz experimental em termos...*/
proc iml;                   /*... das condições experimentais (cexp)*/
/* PROGRAMA DE GERAÇÃO DOS DADOS */
*****igual ao programa auxiliar GERA descrito em A *****

/* PROGRAMA AUXILIAR PARA CONSTRUÇÃO DA MATRIZ DE PLANEJAMENTO DO MODELO
DA VARIÂNCIA */
*****igual ao programa auxiliar PLAN descrito em A *****

/* PROGRAMA AUXILIAR PARA REGRESSÃO USANDO MQO */
*****igual ao programa auxiliar REG descrito em A *****

/* PROGRAMA PARA ESTIMAR OS MODELOS DA MÉDIA E DA VARIÂNCIA */
start ESTIMA;
      /* MÉTODO RQ */
IF MET_S2=0 THEN DO;
***** inclui-se aqui o conteúdo do programa ESTIMA descrito em A *****

      /* MÉTODO S2 */
IF MET_S2=1 THEN DO;
/* Cálculo do número de replicações e das variâncias */
IF ag=0 then do;          /*para quando não for agregado fatores como pseudo-replicações*/
  nrep=j(ncexp,1,.);      /*número de replicações em cada ponto experimental*/
  var=j(ncexp,1,.);       /*variância amostral em cada ponto experimental*/
  j=0;
  do i=1 to ncexp;
    soma=0;
    somaq=0;
    controle=0;
    nrep[i] = 0;
    do while(controle=0);
      j=j+1;
      nrep[i]=nrep[i]+1;
      soma=soma+resposta[j];
      somaq=somaq+resposta[j]**2;
      if j=n_amost then do;
        controle=1;
        goto fim2;
      end;
      if cexp2[j+1] ^= cexp2[j] then controle=1;
    fim2:
    end;
    var[i] = (somaq - soma**2/nrep[i])/(nrep[i]-1);
  end;
END;
if ag = 1 then do;        /*para quando for agregado fatores como pseudo-replicações*/
  varag=j(ncexp,1,.);     /*variância amostral em cada ponto experimental*/
  ncexpv = 2*ncexp;       /*novo número de pontos experimentais*/
  do i=1 to ncexpv;
    ii=i/2;
    i1=i-1;
    if mod(i,2)=0 then do;
      varag[ii] = (var[i1]+var[i])/2; /*agregação das variâncias*/
    end;
  end;
  var=varag;

```

```

end;
/* Modelos para as variâncias */
y = log(var);
x = zz2;
RUN REG;
END;
finish ESTIMA;

/* PROGRAMA PRINCIPAL */
Start PROG;
qq=ncol(zz);
yreal = xx*beta;
vreal = exp(zz*teta);
seed=968974123;
t = j(n_simula,qq,.);
ta=t;
Do simula=1 to n_simula;
run GERA;
/* modelo com variâncias obtidas pelas replicações genuínas */
ncexp=n_cexp;
nfat=n_fat;
fat=fat1;
ag=0;
run PLAN;
run ESTIMA;
t[simula,]=coef;
/* modelo com variâncias obtidas considerando o último fator como replicação adicional */
ncexp=n_cexp/2;
nfat=n_fat-1;
fat=fat2;
ag=1;
run PLAN;
run ESTIMA;
ta[simula,]=coef;
End;
eq = (t-repeat(teta`,n_simula,1))#(t-repeat(teta`,n_simula,1));
eq2 = (ta-repeat(teta`,n_simula,1))#(ta-repeat(teta`,n_simula,1));
eqm=eq[+,]/n_simula;
eqmag=eq2[+,]/n_simula;
efic=eqm/eqmag;
ef[ensaio] = efic[2];
finish PROG;

/* META PROGRAMA PARA EFETUAR A SIMULAÇÃO SOB UM PLANEJAMENTO
EXPERIMENTAL' */

start META;
n_ensaio=32;
ef=j(n_ensaio,1,.);
use planej;
read all var{run a b c d e} into planej;
n_amost = 32;
n_cexp = 16;
n_simula = 2000;
niter=2;
n_fat=4;
yreal=j(n_amost, 1,.);
vreal=j(n_amost, 1,.);
um = j(n_amost,1,1);
use xdad;
read all var {cexp x1 x2 x3 x4};
fat1 = x1 || x2 || x3 || x4;
fat2 = x1 || x2 || x3;
x12=x1#x2;

```

```

x13=x1#x3;
x23=x2#x3;
/*****/
ensaio=1;
met_s2 = 0;
beta={0, 1, 1, 1};
teta={0, .549};
xx = um || x1 || x2 || x3;
zz = um || x1;
run PROG;
/*****/
ensaio=2;
met_s2 = 1;
beta={0, 1, 1, 1};
teta={0, .549};
xx = um || x1 || x2 || x3;
zz = um || x1;
run PROG;
/*****/
ensaio=3;
met_s2 = 0;
beta={0, 1, 1, 1};
teta={0, .896};
xx = um || x1 || x2 || x3;
zz = um || x1;
run PROG;
/*****/
ensaio=4;
met_s2 = 1;
beta={0, 1, 1, 1};
teta={0, .896};
xx = um || x1 || x2 || x3;
zz = um || x1;
run PROG;
/*****/
ensaio=5;
met_s2 = 0;
beta={0, 1, 1, 1, 1};
teta={0, .549};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
run PROG;
/*****/
ensaio=6;
met_s2 = 1;
beta={0, 1, 1, 1, 1};
teta={0, .549};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
run PROG;
/*****/
ensaio=7;
met_s2 = 0;
beta={0, 1, 1, 1, 1};
teta={0, .896};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
run PROG;
/*****/
ensaio=8;
met_s2 = 1;
beta={0, 1, 1, 1, 1};
teta={0, .896};
xx = um || x1 || x2 || x3 || x4;

```

```

zz = um || x1;
run PROG;
/*****/
ensaio=9;
met_s2 = 0;
beta={0, 1, 1, 1};
teta={0, .549, .347, .347};
xx = um || x1 || x2 || x3;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=10;
met_s2 = 1;
beta={0, 1, 1, 1};
teta={0, .549, .347, .347};
xx = um || x1 || x2 || x3;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=11;
met_s2 = 0;
beta={0, 1, 1, 1};
teta={0, .896, .347, .347};
xx = um || x1 || x2 || x3;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=12;
met_s2 = 1;
beta={0, 1, 1, 1};
teta={0, .896, .347, .347};
xx = um || x1 || x2 || x3;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=13;
met_s2 = 0;
beta={0, 1, 1, 1, 1};
teta={0, .549, .347, .347};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=14;
met_s2 = 1;
beta={0, 1, 1, 1, 1};
teta={0, .549, .347, .347};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=15;
met_s2 = 0;
beta={0, 1, 1, 1, 1};
teta={0, .896, .347, .347};
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=16;
met_s2 = 1;
beta={0, 1, 1, 1, 1};
teta={0, .896, .347, .347};
xx = um || x1 || x2 || x3 || x4;

```

```

zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=17;
met_s2 = 0;
beta={0, 1, 1, 1, .5, .5, .5};
teta={0, .549};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1;
run PROG;
/*****/
ensaio=18;
met_s2 = 1;
beta={0, 1, 1, 1, .5, .5, .5};
teta={0, .549};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1;
run PROG;
/*****/
ensaio=19;
met_s2 = 0;
beta={0, 1, 1, 1, .5, .5, .5};
teta={0, .896};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1;
run PROG;
/*****/
ensaio=20;
met_s2 = 1;
beta={0, 1, 1, 1, .5, .5, .5};
teta={0, .896};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1;
run PROG;
/*****/
ensaio=21;
met_s2 = 0;
beta={0, 1, 1, 1, 1, .5, .5, .5};
teta={0, .549};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1;
run PROG;
/*****/
ensaio=22;
met_s2 = 1;
beta={0, 1, 1, 1, 1, .5, .5, .5};
teta={0, .549};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1;
run PROG;
/*****/
ensaio=23;
met_s2 = 0;
beta={0, 1, 1, 1, 1, .5, .5, .5};
teta={0, .896};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1;
run PROG;
/*****/
ensaio=24;
met_s2 = 1;
beta={0, 1, 1, 1, 1, .5, .5, .5};
teta={0, .896};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;

```

```

zz = um || x1;
run PROG;
/*****/
ensaio=25;
met_s2 = 0;
beta={0, 1, 1, 1, .5, .5, .5};
teta={0, .549, .347, .347};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=26;
met_s2 = 1;
beta={0, 1, 1, 1, .5, .5, .5};
teta={0, .549, .347, .347};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=27;
met_s2 = 0;
beta={0, 1, 1, 1, .5, .5, .5};
teta={0, .896, .347, .347};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=28;
met_s2 = 1;
beta={0, 1, 1, 1, .5, .5, .5};
teta={0, .896, .347, .347};
xx = um || x1 || x2 || x3 || x12 || x13 || x23;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=29;
met_s2 = 0;
beta={0, 1, 1, 1, 1, .5, .5, .5};
teta={0, .549, .347, .347};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=30;
met_s2 = 1;
beta={0, 1, 1, 1, 1, .5, .5, .5};
teta={0, .549, .347, .347};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=31;
met_s2 = 0;
beta={0, 1, 1, 1, 1, .5, .5, .5};
teta={0, .896, .347, .347};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;
zz = um || x1 || x2 || x3;
run PROG;
/*****/
ensaio=32;
met_s2 = 1;
beta={0, 1, 1, 1, 1, .5, .5, .5};
teta={0, .896, .347, .347};
xx = um || x1 || x2 || x3 || x4 || x12 || x13 || x23;

```



```

zz = um || x1 || x2 || x3;
run PROG;
/*****/
print planej[format=3.0] ef[format=5.3];
finish META;
RUN META;

```

C - Distribuição $\log(\chi_1 + q)$

Programas computacionais da seção 3.3.1. O primeiro avalia as variâncias da transformada, considerando dados gerados com diferentes variâncias. O segundo avalia a distribuição da transformada usando dados gerados com variância unitária.

*** DISTRIBUIÇÃO DE $\log(S2+q)$, $q = c \cdot S2bar$ com $c=0, 1\%$ e 10% ;**

```

proc iml;
Start gera;
  n=10000;
  seed=351523776;
  normal=0;
  y=j(n,1,.);
  z=j(n,1,.);
  do i=1 to n;
    call rannor(seed, normal);
    y[i] = normal; /*dados normais com média zero e variância um*/
    call rannor(seed, normal);
    z[i] = normal*3; /*dados normais com média zero e variância nove*/
  end;
  y2 = y##2; /*resíduos quadráticos*/
  z2 = z##2;
  u = y2 || z2; /*união dos resíduos quadráticos com diferentes variâncias*/
  umed=u[+]/(2*n); /*média dos resíduos quadráticos*/
  print 'Média global das variâncias:' umed;
  logy2_0=log(y2); /*transformação log. padrão*/
  logz2_0=log(z2);
  logy2_1=log(y2+0.01*umed); /*transformação log. modificada*/
  logz2_1=log(z2+0.01*umed);
  logy2_10=log(y2+0.1*umed);
  logz2_10=log(z2+0.1*umed);
  sai = logy2_0 || logz2_0 || logy2_1 || logz2_1 || logy2_10 || logz2_10; /*cria arquivo*/
  nome_sai = {'logy2_0', 'logz2_0', 'logy2_1', 'logz2_1', 'logy2_10', 'logz2_10'};
  create dad from sai[colname=nome_sai];
  append from sai;
finish GERA;
run gera;
proc means data=dad mean var; /*calcula médias e variâncias das transformadas*/
  var logy2_0 logz2_0 logy2_1 logz2_1 logy2_10 logz2_10;
  title 'Médias e variâncias da distrib. de  $\log(S2+q)$ ,  $q = c \cdot S2bar$  com  $c=0, 1\%$  e  $10\%$ ';
run;

```

*** DISTRIBUIÇÃO DE $\log(Z2+q)$, com $Z \sim N(0,1)$ E $q = 0, 0,01$ e $0,10$;**

```

proc iml;
Start gera;
  n=10000;
  q=0;
  seed=351523776;
  normal=0;
  z=j(n,1,.);
  do i=1 to n;

```

```
call rannor(seed, normal);
  z[i] = normal;
end;
z2 = z##2;
logz2_0 =log(z2);
logz2_1 =log(z2+0.01);
logz2_10 =log(z2+0.1);
sai= logz2_0 || logz2_1 || logz2_10;
nome_sai = {'logz2_0', 'logz2_1','logz2_10'};
create dad from sai[colname=nome_sai];
append from sai;
finish GERA;
run gera;
proc univariate plot data=dad;
  var logz2_0 logz2_1 logz2_10;
  title 'DISTRIBUIÇÃO DE log(Z2+q), com Z~N(0,1) E q = 0, 0,01 e 0,10';
run;
```

D - Avaliação da transformação logarítmica modificada: simulação

Programa computacional do estudo de simulação número 3 (seção 3.3.2).

Inclui os ensaios de 1 a 32. Nos ensaios de 33 a 64, altera-se “data xdad”.

```
** EXPERIMENTO CCD 4 FATORES 31 RUNS ;
* EFEITO DA ADIÇÃO DA CONSTANTE NOS MÉTODOS S2 E RQ;
data planej;
  input run a b c d e f g h;          /*matriz de planejamento da simulação*/
  cards;
1      1      1      1      1      1      1      1      1
2      1      1      1      1      1      -1     -1     -1
3      1      1      1      1      1      -1     1      -1
4      1      1      1      1      -1     -1     1      1
5      1      1      1      -1     1      1      -1     1
6      1      1      1      -1     1      -1     -1     -1
7      1      1      1      -1     -1     1      -1     -1
8      1      1      1      -1     -1     -1     -1     1
9      1      1      -1     1      1      1      -1     1
10     1      1      -1     1      1      -1     -1     -1
11     1      1      -1     1      1      -1     1      -1
12     1      1      -1     1      -1     -1     -1     1
13     1      1      -1     -1     1      1      1      1
14     1      1      -1     -1     1      -1     1      -1
15     1      1      -1     -1     -1     1      1      -1
16     1      1      -1     -1     -1     -1     1      1
17     1      -1     1      1      1      1      -1     -1
18     1      -1     1      1      1      -1     -1     1
19     1      -1     1      1      1      -1     -1     1
20     1      -1     1      1      -1     -1     -1     -1
21     1      -1     1      -1     1      1      1      -1
22     1      -1     1      -1     1      -1     1      1
23     1      -1     1      -1     -1     1      1      1
24     1      -1     1      -1     -1     -1     1      -1
25     1      -1     -1     1      1      1      1      -1
26     1      -1     -1     1      1      -1     1      1
27     1      -1     -1     1      1      -1     1      1
28     1      -1     -1     1      -1     -1     1      -1
29     1      -1     -1     -1     1      1      -1     -1
30     1      -1     -1     -1     -1     1      -1     1
31     1      -1     -1     -1     -1     -1     -1     1
32     1      -1     -1     -1     -1     -1     -1     -1
;
```

```

data xdad;                                /*matriz de planejamento do experimento simulado: CCD 4 fatores*/
input pe a1-a4;
cards;
1      -1      -1      -1      -1
2      -1      -1      -1      1
3      -1      -1      1      -1
4      -1      -1      1      1
5      -1      1      -1      -1
6      -1      1      -1      1
7      -1      1      1      -1
8      -1      1      1      1
9      1      -1      -1      -1
10     1      -1      -1      1
11     1      -1      1      -1
12     1      -1      1      1
13     1      1      -1      -1
14     1      1      -1      1
15     1      1      1      -1
16     1      1      1      1
17     -2      0      0      0
18     2      0      0      0
19     0      -2      0      0
20     0      2      0      0
21     0      0      -2      0
22     0      0      2      0
23     0      0      0      -2
24     0      0      0      2
25     0      0      0      0
26     0      0      0      0
27     0      0      0      0
28     0      0      0      0
29     0      0      0      0
30     0      0      0      0
31     0      0      0      0
;
proc iml;

/* PROGRAMA AUXILIAR PARA FORMAR A MATRIZ DE PLANEJAMENTO DO MODELO DA
MÉDIA */
start EXP;                                /*amplia o número de linhas em função das replicações*/
  n_amost = 31*repl;
  cexp=j(n_amost,1,.);
  x1=cexp; x2=cexp; x3=cexp; x4=cexp;
  do j=1 to ncexp;
    n1=(j-1)*repl+1;
    n2=(j-1)*repl+repl;
    do i=n1 to n2;
      cexp[i]=pe[j];
      x1[i]=a1[j];
      x2[i]=a2[j];
      x3[i]=a3[j];
      x4[i]=a4[j];
    end;
  end;
  x12=x1#x2;
  x13=x1#x3;
  x14=x1#x4;
  x23=x2#x3;
  x24=x2#x4;
  x34=x3#x4;
  x11=x1#x1;
  x22=x2#x2;
  x33=x3#x3;
  x44=x4#x4;

```

```

yreal=j(n_amost, 1,.);
vreal=j(n_amost, 1,.);
um = j(n_amost,1,1);
finish EXP;

/* PROGRAMA DE GERAÇÃO DOS DADOS */
***** programa auxiliar GERA descrito em A ***** /*incluir programa auxiliar GERA*/

/* PROGRAMA AUXILIAR PARA CONSTRUÇÃO DA MATRIZ DE PLANEAMENTO DO MODELO DE
VARIÂNCIA */
start PLANVAR;
zz2 = j(ncexp,qq,.); /*ajusta as linhas em termos do número de pto's experimentais*/
do j=1 to qq;
  zz2[1,j] = zz[1,j];
end;
ii = 1;
do i=2 to n_amost;
  i1=i-1;
  if cexp[i] ^= cexp[i1] then do;
    ii=ii+1;
    do j = 1 to qq;
      zz2[ii,j] = zz[i,j];
    end;
  end;
end;
finish PLANVAR;

/* PROGRAMA AUXILIAR PARA REGRESSAO USANDO MQO */
***** programa auxiliar REG descrito em A *****

/* PROGRAMA PARA ESTIMAR OS MODELOS DA MÉDIA E DA VARIÂNCIA */
start ESTIMA;
  /* MÉTODO S2 */
  IF met_s2=1 then DO;
  /* Cálculo do número de replicações e das variâncias amostrais */
  ***** análogo à parte do programa descrito em B *****

  /* Modelo de regressão das variâncias */
  s2med = var[+]/ncexp; /*média das variâncias amostrais*/
  cte2 = cmed*s2med; /*constante c da transf. log. modificada*/
  if cmed=0 then cte3=0; else cte3 = 0.005*s2med; /*valor mínimo das variâncias*/
  y = log(var+cte2); /*transf. log. modificada*/
  x = zz2;
  RUN REG; /*chama programa auxiliar para a regressão*/
  v2c = exp(yhat)-cte2; /*variâncias preditas - método S2*/
  do i = 1 to ncexp;
    if v2c[i] < cte3 then v2c[i]=cte3; /*fixação do valor mínimo das variâncias preditas*/
  end;
  END;

  /* MÉTODO RQ */
  IF met_s2=0 then DO;
  /* Cálculo do número de replicações */
  ***** análogo à parte do programa descrito em A *****

  /* Modelo inicial da média - MQO */
  y = resposta;
  x = xx;
  RUN REG;
  r = resid/sqrt(1-h); /*resíduos corrigidos*/

  /* Modelo inicial da variância */
  rq = r##2;
  rqm=j(ncexp,1,.); /*cálculo do resíduo quadr. médio de cada ponto experimental*/
  fa=0;
  do i=1 to ncexp;

```

```

soma=0;
do j=1 to nrep[i];
  if i=1 then jj=j;
  else jj=j+fa;
  soma=soma+rq[jj];
end;
rqm[i]=soma/nrep[i];
fa=fa+nrep[i];
end;
rqmed = rqm[+]/ncexp;          /*estimativa do erro (resíduo) quadrático médio*/
cte = cmed*rqmed;              /*definição da constante da transf. log. modificada*/
if cmed=0 then cte3=0; else cte3=0.005*rqmed;
y = log(rqm+cte);
x = zz2;
RUN REG;
vc = exp(yhat) - cte;          /*variâncias preditas - método RQ*/
do i = 1 to ncexp;
  if vc[i] < cte3 then vc[i]=cte3;
end;
/*Processo iterativo*/
DO iter=1 to niter;
/* Modelo da média usando mínimos quadrados ponderados */
va=j(n_amost,1,.);             /*expansão do vetor de variâncias preditas*/
cum = 0;
do i=1 to ncexp;
  if i > 1 then cum = cum + nrep[i-1];
  ii = nrep[i];
  do j=1 to ii;
    va[cum+j] = vc[i];
  end;
end;
w = 1/sqrt(va);                /*raiz dos pesos para transformar as variáveis originais*/
y = resposta # w;
x = xx # w;
RUN REG;
xb=xx*coef;
rnovo= (resposta-xb)/sqrt(1-h); /*resíduos MQG corrigidos*/
rqnovo=rnovo##2;
/* Modelo da variância */
rqm=j(ncexp,1,.);
fa=0;
do i=1 to ncexp;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rqnovo[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;
y = LOG(rqm +cte);
x = zz2;
RUN REG;
vc = exp(yhat)-cte;
do i = 1 to ncexp;
  if vc[i] < cte3 then vc[i]=cte3;
end;
END;
END;
finish ESTIMA;

/* PROGRAMA PRINCIPAL DE CADA ENSAIO */
Start PROG;

```

```

qq=ncol(zz);
yreal = xx*beta;
vreal = g*exp(zz*teta);
t = j(n_simula,1,.); /*vetor que vai guardar as estimativas com transf. log. padrão*/
tc= j(n_simula,ncte,.); /*vetor que vai guardar as estimativas com transf. log. modificada*/
/* estimação com a transformação log padrão */
cmed=0;
seed=1613218064;
do simula=1 to n_simula;
    run GERA;
    run PLANVAR;
    run ESTIMA;
    t[simula,]=coef[2]; /*apenas a estimativa do parâmetro focal*/
end;
tmed = t[+]/n_simula;
sq = (t-tmed)*(t-tmed);
tet=teta[2];
eq = (t-tet)*(t-tet);
eqm = eq[+]/n_simula;
var1[ensaio] = sq[+]/(n_simula-1);
vic1[ensaio] = (tmed-tet)/tet;
/* estimação com a adição de constantes */
cmed=0.005; /*valor inicial de c*/
n1 = (ensaio-1)*ncte + 1;
n2 = (ensaio-1)*ncte + ncte;
do icte=n1 to n2;
    jk=icte-n1+1;
    seed=1613218064;
    do simula=1 to n_simula;
        run GERA;
        run PLANVAR;
        run ESTIMA;
        tc[simula,jk]=coef[2];
    end;
    tcc=tc[,jk];
    tcmed = tcc[+]/n_simula;
    sqc = (tcc-tcmed)*(tcc-tcmed);
    eqc = (tcc-tet)*(tcc-tet);
    eqmc = eqc[+]/n_simula;
    efic[icte] = eqm/eqmc;
    var_c[icte] = sqc[+]/(n_simula-1);
    vic_c[icte] = (tcmed-tet)/tet;
    efcv[icte] = (var1[ensaio]/tmed**2)/(var_c[icte]/tcmed**2);
    ad_cte[icte]=cmed;
    cmed = 2 * cmed; /*constantes c aumentando segundo uma PG de razão 2*/
end;
finish PROG;

/* PROGRAMA PARA EXECUTAR OS DIVERSOS EXPERIMENTOS SIMULADOS, CONFORME A
MATRIZ DE PLANEJAMENTO PLANEJ' */
start META;
    n_ensaio=32; /*número de experimentos simulados com o CCD 4 fat.*/
    ncte = 5; /*número de constantes que serão avaliadas*/
    nresp = ncte*n_ensaio;
    efic = j(nresp,1,.);
    var_c=efic; vic_c=efic; efcv=efic; ad_cte=efic;
    var1=j(n_ensaio,1,.); vic1=var1;
    use planej;
    read ali var{run a b c d e f g h} into planej;
    n_simula = 1000; /*número de amostras simuladas em cada ensaio*/
    ncte = 5;
    niter=2;
    ncexp = 31; /*número de pontos experimentais*/
    use xdad;

```



```

read all var {pe a1 a2 a3 a4};
/*****/ /*definição dos ensaios segundo as condições de PLANEJ*/
ensaio=1;
met_s2 = 0;
repl=4;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={8, 8, 8, 8, 8, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2};
teta={0, .896,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=2;
met_s2 = 1;
repl=4;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={8, 8, 8, 8, 8, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2};
teta={0, .549,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=3;
met_s2 = 1;
repl=4;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={1, 1, 1, 1, 1, .5, .5, .5, .5, .5, .5, .5, .5, .5};
teta={0, .896,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=4;
met_s2 = 0;
repl=4;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={1, 1, 1, 1, 1, .5, .5, .5, .5, .5, .5, .5, .5, .5};
teta={0, .549,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=5;
met_s2 = 0;
repl=2;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={8, 8, 8, 8, 8, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2};

```

```

teta={0, .896,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=6;
met_s2 = 1;
repl=2;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={8, 8, 8, 8, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2};
teta={0, .549,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=7;
met_s2 = 1;
repl=2;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={1, 1, 1, 1, 1, .5, .5, .5, .5, .5, .5, .5, .5, .5};
teta={0, .896,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=8;
met_s2 = 0;
repl=2;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={1, 1, 1, 1, 1, .5, .5, .5, .5, .5, .5, .5, .5, .5};
teta={0, .549,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=9;
met_s2 = 0;
repl=4;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1;
beta={8, 8, 8, 8, 8, 2, 2, 2, 2, 2, 2, 2, 2, 2};
teta={0, .896};
run PROG;
/*****/
ensaio=10;
met_s2 = 1;
repl=4;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1;
beta={8, 8, 8, 8, 8, 2, 2, 2, 2, 2, 2, 2, 2, 2};
teta={0, .549};

```

```

run PROG;
/*****/
ensaio=11;
met_s2 = 1;
repl=4;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1;
beta={1, 1, 1, 1, 1, .5, .5, .5, .5, .5, .5, .5, .5, .5, .5};
teta={0, .896};
run PROG;
/*****/
ensaio=12;
met_s2 = 0;
repl=4;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1;
beta={1, 1, 1, 1, 1, .5, .5, .5, .5, .5, .5, .5, .5, .5, .5};
teta={0, .549};
run PROG;
/*****/
ensaio=13;
met_s2 = 0;
repl=2;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1;
beta={8, 8, 8, 8, 8, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2};
teta={0, .896};
run PROG;
/*****/
ensaio=14;
met_s2 = 1;
repl=2;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1;
beta={8, 8, 8, 8, 8, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2};
teta={0, .549};
run PROG;
/*****/
ensaio=15;
met_s2 = 1;
repl=2;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1;
beta={1, 1, 1, 1, 1, .5, .5, .5, .5, .5, .5, .5, .5, .5, .5};
teta={0, .896};
run PROG;
/*****/
ensaio=16;
met_s2 = 0;

```

```

repl=2;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
zz = um || x1;
beta={1, 1, 1, 1, 1, .5, .5, .5, .5, .5, .5, .5, .5, .5, .5};
teta={0, .549};
run PROG;
/*****/
ensaio=17;
met_s2 = 1;
repl=4;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4 ;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={8, 8, 8, 8, 8};
teta={0, .896,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=18;
met_s2 = 0;
repl=4;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={8, 8, 8, 8, 8};
teta={0, .549,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=19;
met_s2 = 0;
repl=4;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={1, 1, 1, 1, 1};
teta={0, .896,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=20;
met_s2 = 1;
repl=4;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={1, 1, 1, 1, 1};
teta={0, .549,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=21;
met_s2 = 1;
repl=2;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;

```

```

zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={8, 8, 8, 8, 8};
teta={0, .896,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=22;
met_s2 = 0;
repl=2;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={8, 8, 8, 8, 8};
teta={0, .549,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=23;
met_s2 = 0;
repl=2;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={1, 1, 1, 1, 1};
teta={0, .896,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=24;
met_s2 = 1;
repl=2;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1 || x2 || x3 || x4 ||
      x11 || x22 || x33 || x44 || x12 || x13 || x14 || x23 || x24 || x34;
beta={1, 1, 1, 1, 1};
teta={0, .549,.347,.347,.347, .347,.347,.347,.347,.347,.347,.347,.347,.347,.347};
run PROG;
/*****/
ensaio=25;
met_s2 = 1;
repl=4;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
beta={8, 8, 8, 8, 8};
teta={0, .896};
run PROG;
/*****/
ensaio=26;
met_s2 = 0;
repl=4;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
beta={8, 8, 8, 8, 8};
teta={0, .549};
run PROG;
/*****/

```

```

ensaio=27;
met_s2 = 0;
repl=4;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
beta={1, 1, 1, 1, 1};
teta={0, .896};
run PROG;
/*****/
ensaio=28;
met_s2 = 1;
repl=4;
g=1;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
beta={1, 1, 1, 1, 1};
teta={0, .549};
run PROG;
/*****/
ensaio=29;
met_s2 = 1;
repl=2;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
beta={8, 8, 8, 8, 8};
teta={0, .896};
run PROG;
/*****/
ensaio=30;
met_s2 = 0;
repl=2;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
beta={8, 8, 8, 8, 8};
teta={0, .549};
run PROG;
/*****/
ensaio=31;
met_s2 = 0;
repl=2;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
beta={1, 1, 1, 1, 1};
teta={0, .896};
run PROG;
/*****/
ensaio=32;
met_s2 = 1;
repl=2;
g=3;
RUN EXP;
xx = um || x1 || x2 || x3 || x4;
zz = um || x1;
beta={1, 1, 1, 1, 1};
teta={0, .549};

```



```

run PROG;
/*****/
aa=j(nresp,1,.);
col=ncol(planej);
planej2=j(nresp,col,.);
do j=1 to n_ensaio;
  n1 = (j-1)*ncte + 1;
  n2 = (j-1)*ncte + ncte;
  ii=0;
  do i=n1 to n2;
    ii = ii+1;
    aa[i] = ii-3;
    planej2[i,] = planej[j,];
  end;
end;
print planej[format=3.0] var1[format=5.3] vic1[format=5.3];
print planej2[format=3.0] aa[format=4.1] efic[format=5.3] efcv[format=5.3]
      var_c[format=5.3] vic_c[format=5.3];
finish META;
RUN META;

```

E - Agregação de variâncias amostrais e resíduos quadráticos: simulação

Programa computacional básico da seção 3.4.

```

** EXPERIMENTO 2^4 2 repl.;
* COMBINAÇÃO DOS MÉTODOS S2 E RQ;
data xdad;                                /*Matriz de planejamento do experimento simulado*/
input cexp x1-x4;
cards;
  1      -1      -1      -1      -1
  2      -1      -1      -1      1
  3      -1      -1      1      -1
  4      -1      -1      1      1
  5      -1      1      -1      -1
  6      -1      1      -1      1
  7      -1      1      1      -1
  8      -1      1      1      1
  9      1      -1      -1      -1
 10      1      -1      -1      1
 11      1      -1      1      -1
 12      1      -1      1      1
 13      1      1      -1      -1
 14      1      1      -1      1
 15      1      1      1      -1
 16      1      1      1      1
 1      -1      -1      -1      -1
 2      -1      -1      -1      1
 3      -1      -1      1      -1
 4      -1      -1      1      1
 5      -1      1      -1      -1
 6      -1      1      -1      1
 7      -1      1      1      -1
 8      -1      1      1      1
 9      1      -1      -1      -1
10      1      -1      -1      1
11      1      -1      1      -1
12      1      -1      1      1
13      1      1      -1      -1
14      1      1      -1      1
15      1      1      1      -1
16      1      1      1      1
;
proc sort; by cexp;                       /*ordenação em termos dos pontos experimentais*/
proc iml;

```

```

START CONTROLE;      /*definição das principais características do experimento simulado*/
n_amost = 32;         /*número de observações*/
n_fat=4;              /*número de fatores*/
ncexp = 16;           /*número de pontos experimentais ensaiados*/
cmed=0.07;            /*valor de c na transformação log. modificada*/
n_simula = 1000;      /*número de amostras geradas na simulação*/
niter = 2;            /*número de iterações no MQG*/
a=0.01;               /*intervalo do nível descritivo de significância em que os pesos da...*/
b=0.50;               /*...variância amostral e do resíduo quadr. médio são ambos não nulos*/
beta={0,1,1,1,1};    /*coeficientes dos termos principais do modelo das médias */
beta2={1,1,1};        /*coeficientes dos termos de interações do modelo das médias */
beta2=delta*beta2;    /*magnitude das interações*/
betagera= beta // beta2; /*coeficientes do modelo da média usados na simulação*/
teta={0,.693,.549,.347,0}; /*coeficientes do modelo da variância usados na simulação*/
pp=nrow(beta);        /*núm. de parâmetros do modelo da média especificados na estimação*/
qq=nrow(teta);        /*núm. de parâ. do modelo da variância especificados na estimação*/
difuso=0;             /*indicador de quando precisa ponderar variâncias amostrais e resíduos*/
rej0=0;               /*contagem do núm. de vezes que o teste acusa falta de ajuste - MQO*/
rej3=0;               /*contagem do núm. de vezes que o teste acusa falta de ajuste - MQG*/
yreal=j(n_amost, 1,.);
vreal=j(n_amost, 1,.);
um = j(n_amost,1,1);
xx = um || x1 || x2 || x3 || x4;      /*matriz de planejamento do modelo da média para estimação*/
zz = um || x1 || x2 || x3 || x4;      /*matriz de planejamento do modelo da variância*/
x12=x1#x2;
x23=x2#x3;
x34=x3#x4;
xgera = um || x1 || x2 || x3 || x4 || x12 || x23 || x34; /*matriz de planej. da média para geração*/
yreal = xgera*betagera; /*equações da média e variância na geração dos dados*/
vreal=exp(zz*teta);
finish CONTROLE;
/* PROGRAMA DE GERAÇÃO DOS DADOS */
***** igual ao programa auxiliar GERA descrito em A *****

/* PROGRAMA AUXILIAR PARA CONSTRUÇÃO DA MATRIZ DE PLANEAMENTO DO MODELO DE
VARIÂNCIA */
***** igual ao programa auxiliar PLANVAR descrito em D *****

/* PROGRAMA AUXILIAR PARA REGRESSAO USANDO MQO */
start REG;
nobs=nrow(x);
npar=ncol(x);
xxi=inv(x`*x);
coef=xxi*x`*y;          /*coeficientes estimados por MQO*/
yhat=x*coef;            /*valores preditos*/
resid = y-yhat;         /*resíduos*/
h = j(nobs,1,.);
do i=1 to nobs;
  h[i] = x[i,]*xxi*x[i,]; /*vetor para correção dos resíduos*/
end;
if difuso=1 then do;    /*somente quando se referir ao modelo da média*/
  ymed=j(nobs,1,.);    /*calcula a média das obs. em cada ponto experimental*/
  fa=0;
  do i=1 to ncexp;
    soma=0;
    do j=1 to nrep[i];
      if i=1 then jj=j;
      else jj=j+fa;
      soma=soma+y[jj];
    end;
    do j=1 to nrep[i];
      if i=1 then jj=j;
      else jj=j+fa;
      ymed[jj] = soma/nrep[i];
    end;
  end;
end;

```

```

end;
fa=fa+nrep[i];
end;
sse=resid`*resid;           /*soma de quadrados dos erros*/
afast=y-ymed;
sspe=afast`*afast;          /* soma de quadrados do erro puro*/
sslof=sse-sspe;             /* soma de quadrados da falta de ajuste*/
gllof=ncexp-npar;           /*graus de liberdade*/
glpe=nobs-ncexp;
f=(sslof/gllof)/(sspe/glpe); /*estatística F para teste de falta de ajuste*/
ps=1-probf(f,gllof,glpe);   /*nível descritivo de significância*/
if ps>b then qmod=0;         /*construção dos pesos para a medida difusa ...*/
if ps<=a then qmod=1;       /*...de agregação de variâncias amostrais e resíduos quadr.*
if (ps<=b & ps>a) then qmod=(b-ps)/(b-a);
pmod=1-qmod;
end;
Finish REG;
/* PROGRAMA PARA ESTIMAR OS MODELOS DA MÉDIA E DA VARIÂNCIA */
start ESTIMA;
/* CÁLCULO DO NÚMERO DE REPLICAÇÕES DAS VARIÂNCIAS AMOSTRAIS */
var=j(ncexp,1,.);           /*análogo aos anteriores*/
nrep=j(ncexp,1,.);
j=0;
do i=1 to ncexp;
soma=0;
somaq=0;
controle=0;
nrep[i] = 0;
do while(controle=0);
j=j+1;
nrep[i]=nrep[i]+1;
soma=soma+resposta[j];
somaq=somaq+resposta[j]**2;
if j=n_amost then do;
controle=1;
goto fim;
end;
if cexp[j+1] ^= cexp[j] then controle=1;
fim:
end;
var[i] = (somaq - soma**2/nrep[i])/(nrep[i]-1);
end;
/* MODELO INICIAL DE MÉDIA - MQO */
difuso=1;
y = resposta;
x = xx;
RUN REG;
r = resid/sqrt(1-h);
if ps <= 0.05 then rej0=rej0+1;           /*contagem de rejeições da adequação do ajuste*/
/* MÉTODO S2 */
/* MODELO DE VARIÂNCIA */
difuso=0;
s2med = var[+]/ncexp;
cte = cmed*s2med;
cte3 = 0.005*s2med;
y = log(var+cte);
x = zz2;
RUN REG;
coefs2=coef;
/* MÉTODO RQ */
/* MODELO INICIAL DE VARIÂNCIA - RQ */
difuso=0;
rq = r##2;
rqm=j(ncexp,1,.);

```

```

fa=0;
do i=1 to ncexp;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rq[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;
y = log(rqm+cte);
x = zz2;
RUN REG;
vrq = exp(yhat) - cte;
do i = 1 to ncexp;
  if vrq[i] < cte3 then vrq[i]=cte3;
end;
/* PROCESSO ITERATIVO DO ESTIMADOR  RQ */
DO ITER=1 TO NITER;
/* MODELO DE MÉDIA USANDO MÍNIMOS QUADRADOS PONDERADOS - RQ*/
va=j(n_amost,1,.);
cum = 0;
do i=1 to ncexp;
  if i > 1 then cum = cum + nrep[i-1];
  ii = nrep[i];
  do j=1 to ii;
    va[cum+j] = vrq[i];
  end;
end;
w = 1/sqrt(va);
y = resposta # w;
x = xx # w;
RUN REG;
xb=xx*coef;
movov= (resposta-xb)/sqrt(1-h);
rqnovo=rnovo##2;
/* MODELO DE VARIÂNCIA USANDO LOG - RQ*/
rqm=j(ncexp,1,.);
fa=0;
do i=1 to ncexp;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rqnovo[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;
y = LOG(rqm +cte);
x = zz2;
RUN REG;
vrq = exp(yhat)-cte;
do i = 1 to ncexp;
  if vrq[i] < cte3 then vrq[i]=cte3;
end;
if iter=niter then coefrq=coef;
END;
/* MÉTODO MISTO: RQS2 */
/* MODELO INICIAL DE VARIÂNCIA - RQS2 */
difuso=0;
rqv = pmod#rqm + qmod#var; /*medida agregada de variância*/
y = log(rqv+cte);

```

```

x = zz2;
RUN REG;
vrs = exp(yhat) - cte;
do i = 1 to ncexp;
  if vrs[i] < cte3 then vrs[i]=cte3;
end;
/* PROCESSO ITERATIVO DO ESTIMADOR    LOG(p.RQ + q.S2) */
DO ITER=1 TO NITER;
/* MODELOS DE MÉDIAS USANDO MÍNIMOS QUADRADOS PONDERADOS - RQS2*/
difuso=1;
va=j(n_amost,1,.);
cum = 0;
do i=1 to ncexp;
  if i > 1 then cum = cum + nrep[i-1];
  ii = nrep[i];
  do j=1 to ii;
    va[cum+j] = vrs[i];
  end;
end;
w = 1/sqrt(va);
y = resposta # w;
x = xx # w;
RUN REG;
xb=xx*coef;
rnovo= (resposta-xb)/sqrt(1-h);
rqnovo=rnovo##2;
if (iter=niter & ps<=0.05) then rej3=rej3+1;
/* MODELOS PARA AS VARIÂNCIAS - RQS2*/
difuso=0;
fa=0;
do i=1 to ncexp;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rqnovo[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;
rqv = pmod#rqm + qmod#var;
y = LOG(rqv +cte);
x = zz2;
RUN REG;
if iter=niter then coefrs=coef;
vrs= exp(yhat)-cte;
do i = 1 to ncexp;
  if vrs[i] < cte3 then vrs[i]=cte3;
end;
END;
finish ESTIMA;

/* PROGRAMA PRINCIPAL */
Start PROG;
delta=0;          /*introduz interações no modelo das médias - não especificadas na estimação*/
ndelta=8;         /*delta cresce segundo uma P.A. com incrementos de 0,1 unidades, ndelta vezes*/
use xdad;
read ail var {cexp x1 x2 x3 x4};
do idelta=1 to ndelta;
  run CONTROLE;
  if idelta=1 then do;
    print '***** MÉTODO S2_RQ AGREG. *****';
    print ' Condições experimentais';
    print n_amost ncexp n_fat a b cmed;
  end;
end;

```

```

    print beta betagera teta;
end;
print '===== coef. das interações (erro de modelo) =' delta '=====';
seed=1613218064;
ts2 = j(n_simula,qq,.); trq=ts2; trs=ts2;
do simula=1 to n_simula;
    run GERA; /*gera os dados do experimento*/
    run PLANVAR; /*ajusta a matriz de planejamento do modelo da variância*/
    run ESTIMA; /*estima os parâmetros do modelo pelos vários métodos*/
    ts2[simula,]=coefs2; /*guardam-se as estimativas do parâmetro focal*/
    trq[simula,]=coefrq;
    trs[simula,]=coefrs;
end; /*calculam-se a eficiência e outras características*/
eqs2 = (ts2-repeat(teta`,n_simula,1))#(ts2-repeat(teta`,n_simula,1));
eqms2 = eqs2[+,]/n_simula;
eqrq = (trq-repeat(teta`,n_simula,1))#(trq-repeat(teta`,n_simula,1));
eqmrq = eqrq[+,]/n_simula;
eqrs = (trs-repeat(teta`,n_simula,1))#(trs-repeat(teta`,n_simula,1));
eqmrs = eqrs[+,]/n_simula;
efic_rq = eqms2/eqmrq;
efic_rs = eqms2/eqmrs;
rej_mqo = rej0 / n_simula;
rej_mqg = rej3 / n_simula;
print 'Proporção de vezes que o teste inicial detectou falta de ajuste (alfa=0,05)';
print rej_mqo;
print 'Proporção de vezes que o teste final detectou falta de ajuste (alfa=0,05)';
print rej_mqg;
print 'Erro quadrático médio';
print 'S2' eqms2;
print 'RQ' eqmrq;
print 'S2RQ' eqmrs;
print 'Eficiência do RQ em relação ao S2';
print efic_rq;
print 'Eficiência do RQS2 em relação ao S2';
print efic_rs;
delta = delta + 0.1;
end;
finish PROG;
run PROG;

```

F - Correlações entre logaritmos de variâncias amostrais

- * ESTUDO DA CORRELAÇÃO ENTRE O LOG. DE VARIÂNCIAS;
- * DEFINIÇÕES INICIAIS: RO = 0,4 E 2 REPLICAÇÕES (ver final do programa);

```

proc iml;
/* DEFINIÇÕES DOS PARÂMETROS ESTOCÁSTICOS DA SIMULAÇÃO */
start SIG;
    sig = j(2,2,.);
    sig[1,1]=varian1;
    sig[2,2]=varian2;
    sig[1,2] = corr*sqrt(sig[1,1]*sig[2,2]);
    sig[2,1]=sig[1,2];
    avet = eigvec(sig);
    aval = eigval(sig);
    print 'Matriz de covariâncias, autovalores e autovetores';
    print sig aval avet;
finish SIG;

/* PROGRAMA DE GERAÇÃO DOS DADOS */
start GERA;

```



```

cp = j(n_gera,2,-);
s1 = sqrt(aval[1]);
s2 = sqrt(aval[2]);
do i=1 to n_gera;
  cp[i,1] = rannor(0)*s1;
  cp[i,2] = rannor(0)*s2;
end;
erro = cp*inv(avet);
y1 = erro[,1];
y2 = erro[,2];
finish GERA;

/* PROGRAMA AUXILIAR PARA CÁLCULO DA MATRIZ DE CORRELAÇÕES */
Start CORR;
corr = i(kk);
nobs=nrow(mat);
med = mat[,] / nobs;
mat = mat - repeat(med,nobs,1);
do k=1 to kk-1;
  do l=k+1 to kk;
    rkl = mat[,k]#mat[,l];
    num=rkl[+];
    rk2=mat[##,k];
    rl2=mat[##,l];
    corr[k,l] = num/sqrt(rk2*rl2);
    corr[l,k] = corr[k,l];
  end;
end;
Finish CORR;

/* PROGRAMA BÁSICO */
start prog;
n_gera = nrep*n_simul;
nvar=n_simul;
run sig;
run gera;
/* CÁLCULO DAS VARIÂNCIAS AMOSTRAIS */
var1=j(nvar,1,-);
var2=j(nvar,1,-);
j=0;
do i=1 to nvar;
  soma1=0;
  somaq1=0;
  soma2=0;
  somaq2=0;
  do j=1 to nrep;
    jj = j + (i-1)*nrep;
    soma1=soma1 +y1[jj];
    somaq1=somaq1 +y1[jj]**2;
    soma2=soma2 +y2[jj];
    somaq2=somaq2 +y2[jj]**2;
  end;
  var1[i] = (somaq1 - soma1**2/nrep)/(nrep-1);
  var2[i] = (somaq2 - soma2**2/nrep)/(nrep-1);
end;
lvar1=log(var1);
lvar2=log(var2);
mat = var1 ||var2;
run corr;
print 'correlação das variâncias' corr;
mat = lvar1 || lvar2;
run corr;
print 'correlação dos logaritmo das variâncias' corr;
finish prog;

```

```

/* PROGRAMA QUE DEFINE OS DADOS DE ENTRADA */
start entrada;
kk = 2;                /*número de variáveis*/
varian1 = 1;           /*variância da primeira variável*/
varian2 = 1;           /*variância da segunda variável*/
corr = 0.4;            /*correlação entre as variáveis*/
n_simul=10000;         /*número de amostras simuladas*/
nrep = 2;              /*número de replicações consideradas no cálculo das variâncias*/
print '***** PARÂMETROS DE ENTRADA *****';
print nrep corr varian1 varian2 n_simul;
RUN PROG;
finish entrada;
run entrada;

```

G - MQG - SUR: simulação

O programa computacional que segue foi construído para verificar a eficiência dos métodos (a) MQG univariado e (b) SUR incluindo a correção da heterocedasticidade (MQG - SUR). Ambos em relação ao método MQO univariado. O programa está especificado para duas equações, coeficiente de correlação $\rho = 0,40$ e experimento 2^4 , conforme descrito na seção 4.5. A simulação com $\rho = 0,80$ foi feita substituindo-se a linha “ro = 0,40” para “ro = 0,80” no programa auxiliar *controle*. E o experimento CCD foi analisado com adaptações nos dados de entrada “data xdad” e no programa auxiliar *controle*. Para avaliar a eficiência de SUR clássico (supondo homocedasticidade) com relação ao método MQO univariado, executou-se nova rodada do programa, excluindo a correção da heterocedasticidade, mas mantendo-se a mesma sequência de números aleatórios.

- * Sistema de Equações de Regressão Incluindo Modelo de Variância;
- * EXPERIMENTO 2^4 COM 2 REPLICAÇÕES;
- * COMBINAÇÃO DOS MÉTODOS S2 E RQ NO MODELO DA VARIÂNCIA;

```

data xdad;                                /*matriz de planejamento*/
input cexp x1-x4;
cards;
  1      -1      -1      -1      -1
  2      -1      -1      -1      1
  3      -1      -1      1      -1
  4      -1      -1      1      1
  5      -1      1      -1      -1
  6      -1      1      -1      1
  7      -1      1      1      -1
  8      -1      1      1      1
  9      1      -1      -1      -1
 10      1      -1      -1      1
 11      1      -1      1      -1
 12      1      -1      1      1
 13      1      1      -1      -1

```

```

14      1      1      -1      1
15      1      1      1      -1
16      1      1      1      1
1      -1     -1     -1     -1
2      -1     -1     -1      1
3      -1     -1      1     -1
4      -1     -1      1      1
5      -1      1     -1     -1
6      -1      1     -1      1
7      -1      1      1     -1
8      -1      1      1      1
9       1     -1     -1     -1
10      1     -1     -1      1
11      1     -1      1     -1
12      1     -1      1      1
13      1      1     -1     -1
14      1      1      1      1
15      1      1      1     -1
16      1      1      1      1
;
proc sort; by cexp;          /*ordenação dos ensaios em termos dos pontos experimentais*/

proc iml;                    /*inicia-se o procedimento de linguagem matricial*/

/* PROGRAMA COM AS DEFINIÇÕES INICIAIS */
start CONTROLE;
  n_amost = 32;              /*tamanho da amostra*/
  n_fat=4;                   /*número de fatores*/
  ncexp = 16;                /*número de pontos experimentais*/
  cmed=0.06;                 /*valor de c*/
  niter = 2;                 /*número de iterações no MQG*/
  a=0.01;                   /*valores para def. da medida agregada de variância*/
  b=0.50;
  ro = 0.40;                 /*coef. de correlação para a simulação dos dados*/
  beta1={0,1,1,1,1,.5,.5,.5}; /*coef. mod. média da primeira eq.*/
  beta2={0,1,1,1,1};        /*coef. mod. média da segunda eq.*/
  teta1={0,.693,.549,.347};  /*coef. mod. variância da primeira eq.*/
  teta2=teta1;               /*coef. mod. variância da segunda eq.*/
  difuso=0;                  /*indicador para a construção da var. agregada*/
  yreal=j(n_amost, 1,.);
  vreal=j(n_amost, 1,.);
  um = j(n_amost,1,1);
  use xdad;
  read all;
  x12=x1#x2;
  x23=x2#x3;
  x34=x3#x4;                 /*matrizes de planejamento dos modelos*/
  xx1 = um || x1 || x2 || x3 || x4 || x12 || x23 || x34;
  xx2 = um || x1 || x2 || x3;
  zz1 = um || x1 || x2 || x3;
  zz2 = zz1;
  pp = j(kk,1,.);
  qq = j(kk,1,.);
  pp[1]=ncol(xx1);
  pp[2]=ncol(xx2);
  qq[1]=ncol(zz1);
  qq[2]=ncol(zz2);
  yreal1 = xx1*beta1;
  yreal2 = xx2*beta2;
  yreal = yreal1 || yreal2;
  vreal1=exp(zz1*teta1);
  vreal2=exp(zz2*teta2);
  vreal = vreal1 || vreal2;
  print n_fat ncexp n_amost ro;
  print beta1 beta2 teta1 teta2;
finish CONTROLE;

```

```
/* DEFINIÇÕES DOS PARÂMETROS ESTOCÁSTICOS DA SIMULAÇÃO */
```

```
start SIG;
  sig = j(kk,kk,.);
  corrél = j(kk,kk,.);
  sig[1,1]=1;           /*variâncias na simulação*/
  sig[2,2]=1;
  corrél[1,2]= ro;       /*correlação na simulação*/
  do k=1 to kk-1;        /*operador para transf. vetores indep. em vetores correlacionados*/
    do l=k+1 to kk;
      sig[k,l] = corrél[k,l]*sqrt(sig[k,k]*sig[l,l]);
      sig[l,k] = sig[k,l];
    end;
  end;
  avet = eigvec(sig);
  aval = eigval(sig);
  trans=sqrt(diag(aval))*inv(avet);
finish SIG;
```

```
/* PROGRAMA DE GERAÇÃO DOS DADOS */
```

```
start GERA;
  nor = j(n_amost,kk,.);
  normal=0;
  do i=1 to n_amost;
    do k=1 to kk;
      call rannor(seed, normal);
      nor[i,k] = normal;
    end;
  end;
  e = nor*trans;          /*transforma num vetor normal correlacionado*/
  erro = sqrt(vreal) # e; /*simula os termos de erro*/
  m_resp = yreal + erro;  /*simula as respostas*/
finish GERA;
```

```
/* PROGRAMA AUXILIAR PARA CONSTRUÇÃO DA MATRIZ DE PLANEAMENTO DO MODELO DE VARIÂNCIA */
```

```
start PLANVAR;
  zb = j(ncexp,nparv,.); /*ajusta a dimensão da matriz de plan. do mod. da variância*/
  do j=1 to nparv;
    zb[1,j] = zz[1,j];
  end;
  ii = 1;
  do i=2 to n_amost;
    i1=i-1;
    if cexp[i] ^= cexp[i1] then do;
      ii=ii+1;
      do j = 1 to nparv;
        zb[ii,j] = zz[i,j];
      end;
    end;
  end;
finish PLANVAR;
```

```
/* PROGRAMA AUXILIAR PARA REGRESSAO USANDO MQO ou MQG*/
```

```
start REG;
  nobs=nrow(x);
  npar=ncol(x);
  xxi=inv(x`*x);
  coef=xxi*x`*y;
  yhat=x*coef;
  resid = y-yhat;
  h = j(nobs,1,.);
  do i=1 to nobs;
    h[i] = x[i,]*xxi*x[i,]`;
```

```

end;
if difuso=1 then do;
  ymed=j(nobs,1,.);
  fa=0;
  do i=1 to ncexp;
    soma=0;
    do j=1 to nrep[i];
      if i=1 then jj=j;
      else jj=j+fa;
      soma=soma+y[jj];
    end;
    do j=1 to nrep[i];
      if i=1 then jj=j;
      else jj=j+fa;
      ymed[jj] = soma/nrep[i];
    end;
    fa=fa+nrep[i];
  end;
  sse=resid`*resid;
  afast=y-ymed;
  sspe=afast`*afast;
  sslof=sse-sspe;
  gllof=ncexp-npar;
  glpe=nobs-ncexp;
  f=(sslof/gllof)/(sspe/glpe);
  ps=1-probf(f,gllof,glpe);
  if ps>b then qmod=0;
  if ps<=a then qmod=1;
  if (ps<=b & ps>a) then qmod=(b-ps)/(b-a);
  pmod=1-qmod;
end;
Finish REG;

/* PROGRAMA AUXILIAR PARA MÍNIMOS QUADRADOS GENERALIZADOS EM SISTEMAS DE
EQUAÇÕES*/
start SUR;
  npar = ncol(x);
  nt = nrow(x);
  dfe = nt - npar;
  yt = y # w;
  xt = x # w;
/*transformações para a correção da heterocedasticidade*/
/* Obtenção do vetor de parâmetros (bb) */
  is = inv(ss);
  isa = is @ i(n_amostr);
  part1 = inv(xt`*isa*xt);
  bb = part1 * xt`*isa*yt;
/*inversão da matriz de covariâncias (ou correlações)*/
/*estimativa dos coeficientes do sistema*/
Finish SUR;

/* PROGRAMA AUXILIAR PARA CÁLCULO DA MATRIZ DE CORRELAÇÕES */
Start CORR;
  corr = i(kk);
  nobs=nrow(mat);
  med = mat[,] / nobs;
  mat = mat - repeat(med,nobs,1);
  do k=1 to kk-1;
    do l=k+1 to kk;
      rkl = mat[,k]#mat[,l];
      num=rkl[+];
      rk2=mat[##,k];
      rl2=mat[##,l];
      corr[k,l] = num/sqrt(rk2*rl2);
      corr[l,k] = corr[k,l];
    end;
  end;
end;

```

Finish CORR;

/* PROGRAMA PARA ESTIMAR OS MODELOS DAS MÉDIA E DAS VARIÂNCIA */

start ESTIMA;

xx = block(xx1, xx2); /*montagem das matrizes de planejamento ampliadas*/

zz = block(zz1b, zz2b);

/* CÁLCULO DO NÚMERO DE REPLICAÇÕES E DAS VARIÂNCIAS AMOSTRAIS */

nn = n_amost*kk;

m_var=j(ncexp, kk,.);

nrep=j(ncexp, 1, 0);

do k = 1 to kk;

j = 0;

do i=1 to ncexp;

soma=0;

somaq=0;

controle=0;

if k = 1 then nrep[i] = 0;

do while(controle=0);

j=j+1;

if k=1 then nrep[i]=nrep[i]+1;

soma=soma+m_resp[j, k];

somaq=somaq+m_resp[j, k]**2;

if j=n_amost then do;

controle=1;

goto fim;

end;

if cexp[j+1] ^= cexp[j] then controle=1;

fim:

end;

m_var[i, k] = (somaq - soma**2/nrep[i])/(nrep[i]-1);

end;

end;

/* MODELOS UNIVARIADOS PARA AS MÉDIAS */

difuso=1;

p=0;

q=0;

r = j(n_amost, kk,.);

pmodel=j(kk, 1,.);

qmodel=j(kk, 1,.);

do k=1 to kk;

/*separação das partes relativas a cada equação*/

l1 = p+1;

l2 = l1 + n_amost-1;

c1 = q+1;

c2 = c1 + pp[k]-1;

y = m_resp[, k];

x = xx[l1:l2, c1:c2];

RUN REG;

/*realização da regressão*/

r[, k] = resid/sqrt(1-h);

/*correção dos resíduos*/

p = p + n_amost;

q = q + pp[k];

pmodel[k] = pmod;

qmodel[k] = qmod;

if k = 1 then coef1[simula,] = coef;

/*armazenamento das estimativas*/

if k = 2 then coef2[simula,] = coef;

end;

/* MODELOS UNIVARIADOS PARA AS VARIÂNCIAS, COM TRANSF. LOG. */

difuso=0;

p=0;

q=0;

v=j(ncexp, kk,.);

cte=j(kk, 1,.);

cte3=j(kk, 1,.);

do k=1 to kk;

/*cálculo do res. quadr. médio de cada ponto experimental e de cada eq.*/

rq = r[, k]**2;

```

rqm=j(ncexp,1,.);
fa=0;
do i=1 to ncexp;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rq[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;
rqv = pmodel[k]*rqm + qmodel[k]*m_var[,k];      /*medida agregada de variância*/
rqvmed=rqv[+]/ncexp;
cte[k] = cmed*rqvmed;
cte3[k] = 0.005*rqvmed;
l1 = p+1;
l2 = l1 + ncexp-1;
c1 = q+1;
c2 = c1 + qq[k]-1;
y = log(rqv+cte[k]);
x = zz[l1:l2 , c1:c2];
RUN REG;                                          /*realização da regressão para cada equação*/
v[,k] = exp(yhat) - cte[k];                    /*variâncias preditas*/
p = p+ncexp;
q = q + qq[k];
do i=1 to ncexp;
  if v[i,k] < cte3[k] then v[i,k]=cte3[k];
end;
end;
/* EXPANSÃO DA MATRIZ DE VARIÂNCIAS PREDITAS */
va=j(n_amost,kk,.);
do k=1 to kk;
  cum = 0;
  do i=1 to ncexp;
    if i > 1 then cum = cum + nrep[i-1];
    ii = nrep[i];
    do j=1 to ii;
      va[cum+j,k] = v[i,k];
    end;
  end;
end;
/* MODELOS UNIVARIADOS PARA AS MÉDIAS USANDO MQG */
difuso=1;
rhomo = j(n_amost,kk,.);
p=0;
q=0;
do k=1 to kk;
  l1 = p+1;
  l2 = l1 + n_amost-1;
  c1 = q+1;
  c2 = c1 + pp[k]-1;
  yst = m_resp[,k];
  xst = xx[l1:l2 , c1:c2];
  w = 1/sqrt(va[,k]);                          /*raiz dos pesos*/
  y = m_resp[,k] # w;                          /*transf. para MQG*/
  x = xx[l1:l2 , c1:c2] # w;
  RUN REG;
  if k=1 then bb = coef ; else bb = bb // coef;
  rhomo[k] = resid/sqrt(1-h);                  /*resíduos corrigidos com as var. transformadas (w)*/
  p = p+n_amost;
  q = q + pp[k];
  pmodel[k] = pmod;
  qmodel[k] = qmod;

```



```

    if k = 1 then coefmqg1[simula,] = coef;          /*armazenamento das estimativas*/
    if k = 2 then coefmqg2[simula,] = coef;
end;
difuso=0;
/**** PREPARAÇÃO DO PROCESSO MULTIVARIADO *****/
/* MATRIZ DE CORRELAÇÕES */
mat = rhomo;
RUN CORR;
/* TRANSFORMAÇÃO DA MATRIZ DE VAR. DEP. EM VETORES */
p = 0;
do k=1 to kk;
    if k=1 then yy = m_resp[, 1]; else yy = yy // m_resp[,k];
end;
DO ITER=1 TO NITER;
/*** ESTIMATIVA MULTIVARIADA DAS MEDIAS, USANDO O PROGRAMA AUXILIAR SUR */
vv = j(nn,1,.);
do k=1 to kk;          /*vetor ampliado de variâncias preditas p/ correção da heteroced.*/
    if k=1 then vv=va[, 1]; else vv = vv // va[,k];
end;
x = xx;
y = yy;
w = 1/sqrt(vv);          /*pesos para correção da heteroced. - MQG*/
ss = corr;          /*correlação estimada dos resíduos*/
RUN SUR;
p = 0;
q = 0;
do k=1 to kk;
    l1 = p + 1;
    l2 = l1 + n_amost - 1;
    c1 = q + 1;
    c2 = c1 + pp[k] - 1;
    coef = bb[c1:c2];
    p = p+n_amost;
    q = q + pp[k];
    if iter=niter then do;
        if k = 1 then coefs1[simula,] = coef;          /*armazenamento dos coeficientes por equação*/
        if k = 2 then coefs2[simula,] = coef;
    end;
end;
/* CALCULO DOS RESÍDUOS HOMOCEDÁSTICOS */
hh = j(nn,1,.);
rr = yy - xx*bb;
rrhomo = yt - xt*bb;
do i=1 to nn;
    hh[i] = xt[i,]*part1*xt[i,];
    rr[i] = rr[i]/sqrt(1-hh[i]);          /*resíduos corrigidos pelos g.l.*/
    rrhomo[i] = rrhomo[i]/sqrt(1-hh[i]);          /*resíduos corrigidos pelos g.l. e pela heteroced.*/
end;
p = 0;
do k=1 to kk;          /*transf. do vetor de resíduos em matriz (por eq.)*/
    do i=1 to n_amost;
        rhomo[i,k] = rrhomo[p+i];
        r[i,k] = rr[p+i];
    end;
    p = p + n_amost;
end;
mat = rhomo;
run CORR;
if iter=niter then do;
    cor[simula, 1] = corr[1,2];
end;
/* MODELOS DE VARIÂNCIA COM OS NOVOS RESÍDUOS - ESTIMATIVAS UNIVARIADAS */
difuso=0;
p=0;

```

```

q=0;
do k=1 to kk;
  rq = r[,k]##2;
  rqm=j(ncexp, 1,.);
  fa=0;
  do i=1 to ncexp;
    soma=0;
    do j=1 to nrep[i];
      if i=1 then jj=j;
      else jj=j+fa;
      soma=soma+rq[jj];
    end;
    rqm[i]=soma/nrep[i];
    fa=fa+nrep[i];
  end;
  rqv = pmodel[k]*rqm + qmodel[k]*m_var[,k];
  rqvmed=rqv[+]/ncexp;
  l1 = p+1;
  l2 = l1 + ncexp-1;
  c1 = q+1;
  c2 = c1 + qq[k]-1;
  y = log(rqv+cte[k]);
  x = zz[l1:i2 , c1:c2];
  RUN REG;
  v[,k] = exp(yhat) - cte[k];
  p = p+ncexp;
  q = q + qq[k];
  do i=1 to ncexp;
    if v[i,k] < cte3[k] then v[i,k]=cte3[k];
  end;
end;
/* EXPANSÃO DA MATRIZ DE VARIÂNCIAS PREDITAS */
do k=1 to kk;
  cum = 0;
  do i=1 to ncexp;
    if i > 1 then cum = cum + nrep[i-1];
    ii = nrep[i];
    do j=1 to ii;
      va[cum+j,k] = v[i,k];
    end;
  end;
end;
END;
finish ESTIMA;

/* PROGRAMA PRINCIPAL DA SIMULAÇÃO*/
Start PROG;
  n_simula=1000;          /*número de amostras simuladas*/
  kk=2;                  /*núm. de equações*/
  pp1=8;                 /*núm. de parâmetros da eq. 1 do sistema das médias*/
  pp2=4;                 /*núm. de parâmetros da eq. 2 do sistema das médias*/
  seed=1613218064;       /*semente inicial da geração de números aleatórios*/
  coef1=j(n_simula, pp1,.); /*def. dos vetores onde se vai armazenar as estimativas*/
  coef2=j(n_simula, pp2,.);
  coefmqg1=j(n_simula, pp1,.);
  coefmqg2=j(n_simula, pp2,.);
  coefs1=j(n_simula, pp1,.);
  coefs2=j(n_simula, pp2,.);
  cor = j(n_simula,kk,.);
  run CONTROLE;          /*criação dos dados iniciais*/
  run SIG;               /*cria o transformador de vetores univariados em multivariados (correlacionados)*/
  do simula = 1 to n_simula;
    run GERA;            /*gera a amostra*/
    zz=zz1;
  end;

```

```

nparv=ncol(zz);
run PLANVAR;          /*ajusta a matriz de planej. da variância, por equação*/
zz1b = zb;
zz=zz2;
nparv=ncol(zz);
run PLANVAR;
zz2b = zb;
run ESTIMA;           /*procede a estimação pelos vários métodos*/
end;
corr12 = cor[+,1]/n_simula;          /*construção das estatísticas das estimativas*/
print 'Valor esperado das correlações';
print corr12;
eq1 = (coef1-repeat(beta1`,n_simula,1))#(coef1-repeat(beta1`,n_simula,1));
eq2 = (coef2-repeat(beta2`,n_simula,1))#(coef2-repeat(beta2`,n_simula,1));
eqg1 = (coefmqg1-repeat(beta1`,n_simula,1))#(coefmqg1-repeat(beta1`,n_simula,1));
eqg2 = (coefmqg2-repeat(beta2`,n_simula,1))#(coefmqg2-repeat(beta2`,n_simula,1));
eqs1 = (coefs1-repeat(beta1`,n_simula,1))#(coefs1-repeat(beta1`,n_simula,1));
eqs2 = (coefs2-repeat(beta2`,n_simula,1))#(coefs2-repeat(beta2`,n_simula,1));
eqm1=eq1[+,]/n_simula;
eqm2=eq2[+,]/n_simula;
eqmg1=eqg1[+,]/n_simula;
eqmg2=eqg2[+,]/n_simula;
eqms1=eqs1[+,]/n_simula;
eqms2=eqs2[+,]/n_simula;
efmqg1 = eqm1/eqmg1;
efmqg2 = eqm2/eqmg2;
efsur1 = eqm1/eqms1;
efsur2 = eqm2/eqms2;
print 'Eficiência do MQG em relação ao MQO, por equação e por coeficiente';
print efmqg1;
print efmqg2;
print 'Eficiência do MQG-SUR em relação ao MQO, por equação e por coeficiente';
print efsur1;
print efsur2;
Finish PROG;
run PROG;

```

H - Estudo de caso nº 1

Programa computacional do exemplo da moldagem por injeção. Envolve o algoritmo MQG univariado, com o modelo da variância agregando resíduos quadráticos de efeitos de dispersão não significativos. O programa é específico para os modelos do presente estudo.

- ** AGREGAÇÃO DE RESÍDUOS QUADRÁTICOS PARA O MODELO DA VARIÂNCIA;
- ** Estudo de caso n. 1 - Montgomery (1997, p. 391);
- ** Efeitos no modelo da média: x1, x2 e x1*x2;
- ** Efeitos no modelo da variância: x3;

```

data dados;
input cexp x1-x6 resposta;
cards;

```

1	-1	-1	-1	-1	-1	-1	6
2	1	-1	-1	-1	1	-1	10
3	-1	1	-1	-1	1	1	32
4	1	1	-1	-1	-1	1	60
5	-1	-1	1	-1	1	1	4

6	1	-1	1	-1	-1	1	15
7	-1	1	1	-1	-1	-1	26
8	1	1	1	-1	1	-1	60
9	-1	-1	-1	1	-1	1	8
10	1	-1	-1	1	1	1	12
11	-1	1	-1	1	1	-1	34
12	1	1	-1	1	-1	-1	60
13	-1	-1	1	1	1	-1	16
14	1	-1	1	1	-1	-1	5
15	-1	1	1	1	-1	1	37
16	1	1	1	1	1	1	52

;** Ordenação dos dados por nível de x3 (fator de dispersão);

proc sort; by x3;

proc iml; /*procedimento de linguagem matricial*/

/* ESPECIFICAÇÕES INICIAIS */

start INICIO;

n_amost=16; /*número total de observações*/

ncexp=16; /*número de pontos experimentais ensaiados*/

niter=2; /*número de iterações no MQG*/

use dados;

read all var {cexp x1 x2 x3 x4 x5 x6 resposta};

x12=x1#x2; /*definição das interações*/

um = j(n_amost,1,1);

xx = um || x1 || x2 || x12; /*matriz de planejamneto do modelo da média*/

zz2 = {1 -1, 1 1}; /*matriz de planejamneto do modelo da variância*/

ncexp2=2; /*número de pontos experimentais no modelo da variância*/

cexp2 = j(n_amost,1,.);

nrep=j(ncexp2,1,.);

nrep[1]=8; /*número de agregações no modelo da variância*/

nrep[2]=8;

do i = 1 to n_amost;

if x3[i] = -1 then cexp2[i] = 1;

else cexp2[i] = 2; /*número de pontos exper. considerados no modelo da

variância*/

end;

Finish INICIO;

/* PROGRAMA AUXILIAR PARA REGRESSAO USANDO MQO */

start REG;

nobs=nrow(x);

npar=ncol(x);

xxi=inv(x' *x);

coef=xxi*x' *y; /*estimacão pontual dos coeficientes*/

yhat=x*coef;

resid = y-yhat; /*resíduos brutos*/

h = j(nobs,1,.);

do i=1 to nobs;

h[i] = x[i,]*xxi*x[i,]; /*vetor para correção dos resíduos (perdas de g.l.)*/

end;

Finish REG;

/* PROGRAMA PARA ESTIMAR OS MODELOS DA MÉDIA E DA VARIÂNCIA */

start ESTIMA;

/* MODELO INICIAL PARA AS MÉDIAS - MQO */

y = resposta;

x = xx;

RUN REG;

r = resid/sqrt(1-h); /*correção dos resíduos pela perda de g.l.*/

print 'Coeficientes do modelo da média - MQO' coef;

```

/* MODELO INICIAL PARA AS VARIÂNCIAS */
rq = r##2; /*resíduos quadráticos*/
rqm=j(ncexp2,1,.); /*média dos resíduos quadráticos em cada ponto
exper.*/
fa=0;
do i=1 to ncexp2;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rq[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;
y = log(rqm); /*logaritmo dos resíduos quadráticos médios*/
x = zz2;
RUN REG;
eqm = rq[+]/n_amost; /*estimativa do erro quadrático médio - mod. var.*/
coef[1]=log(eqm); /*alteração da estimativa do termo constante*/
v = exp(yhat); /*predição das variâncias*/
print 'Resíduo quadrático médio em cada nível de x3' rqm;
print 'Coeficientes do modelo inicial da variância' coef;

/* PROCESSO ITERATIVO - MQG*/
do iter=1 to niter;

/* MODELO PARA AS MÉDIAS USANDO MÍNIMOS QUADRADOS PONDERADOS - RQ*/
va=j(n_amost,1,.); /*expansão do vetor de variâncias preditas*/
cum = 0;
do i=1 to ncexp2;
  if i > 1 then cum = cum + nrep[i-1];
  ii = nrep[i];
  do j=1 to ii;
    va[cum+j] = v[i];
  end;
end;
w = 1/sqrt(va); /*sistema de pesos do MQG*/
y = resposta # w;
x = xx # w;
RUN REG;
xb=xx*coef;
rnovo= (resposta-xb)/sqrt(1-h); /*novos resíduos corrigidos*/
rqnovo=rnovo##2;
if iter=niter then print 'Coeficientes do modelo da média - MQG' coef;

/* MODELO PARA AS VARIÂNCIAS */
rqm=j(ncexp2,1,.);
fa=0;
do i=1 to ncexp2;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rqnovo[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;
y = LOG(rqm);
x = zz2;
RUN REG;
eqm = rqnovo[+]/n_amost;
coef[1]=log(eqm);

```

```

v= exp(yhat);
if iter=niter then print 'Coeficientes do modelo final da variância' coef;
end;
finish ESTIMA;

start PROG;
run INICIO;
run ESTIMA;
finish PROG;

run PROG;

```

I - Estudo de caso nº 2

Programa computacional do exemplo da catapulta. Envolve o algoritmo MQG univariado, com o modelo da variância incorporando a transformação logarítmica modificada e agregação de variâncias amostrais com resíduos quadráticos médios.

```

** PROGRAMA PARA ESTIMAR OS PARÂMETROS DA EQ. DA MÉDIA E DA VARIÂNCIA;
** DADOS DE LUNER (1994);
** USO DE MQG, TRANSFORMAÇÃO LOG MODIFICADA E COMBINAÇÃO DOS MÉTODOS S2 E
RQ;

```

```

data dados;                                /*entrada dos dados*/
input cexp trat x1 x2 x3 resposta;
cards;
1      1      -1.00000      -1.00000      -1.00000      39
2      2      -1.00000      -1.00000      1.00000      80
3      3      -1.00000      1.00000      -1.00000      52
4      4      -1.00000      1.00000      1.00000      97
5      5      1.00000      -1.00000      -1.00000      60
6      6      1.00000      -1.00000      1.00000      113
7      7      1.00000      1.00000      -1.00000      78
8      8      1.00000      1.00000      1.00000      130
9      9      -1.68179      0.00000      0.00000      59
10     10     1.68179      0.00000      0.00000      115
11     11     0.00000      -1.68179      0.00000      50
12     12     0.00000      1.68179      0.00000      88
13     13     0.00000      0.00000      -1.68179      54
14     14     0.00000      0.00000      1.68179      122
15     15     0.00000      0.00000      0.00000      87
16     15     0.00000      0.00000      0.00000      86
17     15     0.00000      0.00000      0.00000      88
18     15     0.00000      0.00000      0.00000      89
19     15     0.00000      0.00000      0.00000      86
20     15     0.00000      0.00000      0.00000      88
1      1      -1.00000      -1.00000      -1.00000      34
2      2      -1.00000      -1.00000      1.00000      71
3      3      -1.00000      1.00000      -1.00000      44
4      4      -1.00000      1.00000      1.00000      68
5      5      1.00000      -1.00000      -1.00000      53
6      6      1.00000      -1.00000      1.00000      104
7      7      1.00000      1.00000      -1.00000      64
8      8      1.00000      1.00000      1.00000      79
9      9      -1.68179      0.00000      0.00000      51
10     10     1.68179      0.00000      0.00000      102
11     11     0.00000      -1.68179      0.00000      43
12     12     0.00000      1.68179      0.00000      49

```

```

13      13      0.00000      0.00000      -1.68179      50
14      14      0.00000      0.00000      1.68179      109
15      15      0.00000      0.00000      0.00000      78
16      15      0.00000      0.00000      0.00000      79
17      15      0.00000      0.00000      0.00000      81
18      15      0.00000      0.00000      0.00000      82
19      15      0.00000      0.00000      0.00000      79
20      15      0.00000      0.00000      0.00000      79
1       1      -1.00000      -1.00000      -1.00000      42
2       2      -1.00000      -1.00000      1.00000      91
3       3      -1.00000      1.00000      -1.00000      45
4       4      -1.00000      1.00000      1.00000      60
5       5      1.00000      -1.00000      -1.00000      68
6       6      1.00000      -1.00000      1.00000      127
7       7      1.00000      1.00000      -1.00000      65
8       8      1.00000      1.00000      1.00000      75
9       9      -1.68179      0.00000      0.00000      60
10      10      1.68179      0.00000      0.00000      117
11      11      0.00000      -1.68179      0.00000      57
12      12      0.00000      1.68179      0.00000      43
13      13      0.00000      0.00000      -1.68179      60
14      14      0.00000      0.00000      1.68179      119
15      15      0.00000      0.00000      0.00000      89
16      15      0.00000      0.00000      0.00000      85
17      15      0.00000      0.00000      0.00000      87
18      15      0.00000      0.00000      0.00000      87
19      15      0.00000      0.00000      0.00000      88
20      15      0.00000      0.00000      0.00000      90
;
proc sort; by cexp;                /*ordenação dos dados em termos dos pontos experimentais*/
proc iml;                          /*início do procedimento de linguagem matricial*/
START CONTROLE;                  /*parâmetros de entrada*/
  n_fat=3;                        /*número de fatores*/
  ncexp = 20;                     /*número de variâncias a serem calculadas*/
  ntrat = 15;                     /*número de tratamentos (pontos experimentais)*/
  cmed=0.08;
  niter = 2;                      /*número de iterações no processo de estimação*/
  a=0.01;                         /*valores para o cálculo da medida difusa de adequação...*/
  b=0.50;                         /* do modelo da média*/
  difuso=0;
  use dados;
  read all var {cexp trat x1 x2 x3 resposta};
  n_amost = nrow(resposta);       /*tamanho da amostra*/
  x11=x1#x1;
  x22=x2#x2;
  x33=x3#x3;
  x12=x1#x2;
  x13=x1#x3;
  x23=x2#x3;
  um = j(n_amost,1,1);           /*matrizes de planejamentos*/
  xx = um || x1 || x2 || x3 || x11 || x22 || x33 || x12 || x13 || x23; /*modelo da média*/
  zz = um || x1 || x2 || x3 || x11 || x22 || x33 || x12 || x13 || x23; /*modelo da variância*/
  nomesX = {cte, x1, x2, x3, x11, x22, x33, x12, x13, x23};
  nomesZ = {cte, x1, x2, x3, x11, x22, x33, x12, x13, x23};
  pp=ncol(xx);
  qq=ncol(zz);
finish CONTROLE;

/* PROGRAMA AUXILIAR PARA CONSTRUÇÃO DA MATRIZ DE PLANEAMENTO DO MODELO DA
VARIÂNCIA */
start PLANVAR;
  zz2 = j(ncexp,qq,.);           /*ajusta a dimensão da matriz de planejamento: núm. de linhas = ...*/
  do j=1 to qq;                  /*... = núm de pontos experimentais = ncexp*/
    zz2[1,j] = zz[1,j];

```



```

end;
ii = 1;
do i=2 to n_amost;
  i1=i-1;
  if cexp[i] ^= cexp[i1] then do;
    ii=ii+1;
    do j = 1 to qq;
      zz2[ii,j] = zz[i,j];
    end;
  end;
end;
end;
finish PLANVAR;

/* PROGRAMA AUXILIAR PARA REGRESSAO USANDO MQO */
start REG;
nobs=nrow(x);
npar=ncol(x);
xxi=inv(x' * x);
coef=xxi*x'*y;          /*estimativa dos coeficientes*/
yhat=x*coef;            /*valores preditos*/
resid = y-yhat;         /*resíduos*/
h = j(nobs,1,.);
do i=1 to nobs;
  h[i] = x[i,]*xxi*x[i,]; /*vetor para correção dos resíduos*/
end;
ybar=y[+]/nobs;
dt=y-ybar;
sqt=dt*dt;              /*soma de quadrados total*/
sqe=resid*resid;        /*soma de quadrados dos erros*/
sqr=sqt-sqe;            /*soma de quadrados da regressão*/
glt=nobs-1;
gle=nobs-npar;
glr=npar-1;
qme=sqe/gle;
qmr=sqr/gl;
r2=sqr/sqt;             /*coef. de determinação*/
f_mod = qmr/qme;        /*estatística F para teste do modelo*/
p_mod = 1-probf(f_mod,glr,gle); /*nível descritivo de significância*/
covb=xxi#qme;
Err_padr=sqrt(vecdiag(covb)); /*erro padrão das estimativas*/
t=coef/Err_padr;        /*estatística t para o teste de coeficientes*/
probab = 1-probf(t#t,1,gle); /*nível descritivo de significância*/
if difuso=1 then do;    /*preparação para o teste de falta de ajuste*/
  ymed=j(nobs,1,.);    /*cálculo da média em cada ponto experimental*/
  fa=0;
  do i=1 to ntrat;
    soma=0;
    do j=1 to nrep2[i];
      if i=1 then jj=j;
      else jj=j+fa;
      soma=soma+y[jj];
    end;
    do j=1 to nrep2[i];
      if i=1 then jj=j;
      else jj=j+fa;
      ymed[jj] = soma/nrep2[i];
    end;
    fa=fa+nrep2[i];
  end;
  afast=y-ymed;
  sqpe=afast*afast;
  sqfalta=sqe-sqpe;
  glfalta=ntrat-npar;
  glpe=nobs-ntrat;

```

```

f_falta=(sqfalta/gl falta)/(sqpe/glpe);      /*estatística F para o teste de falta de ajuste*/
p_falta=1-probf(f_falta,gl falta,glpe);
if p_falta>b then qmod=0;                      /*pesos para e.q.m. e s2*/
if p_falta<=a then qmod=1;
if (p_falta<=b & p_falta>a) then qmod=(b-p_falta)/(b-a);
pmod=1-qmod;
end;
Finish REG;

/* PROGRAMA PARA ESTIMAR OS MODELOS DA MÉDIA E DA VARIÂNCIA */
start ESTIMA;
run CONTROLE;                                /*chama programa para ter os dados básicos*/
run PLANVAR;                                  /*chama programa para corrigir dimensão da matriz de plan.*/
/* CÁLCULO DO NÚMERO DE REPLICAÇÕES E DAS VARIÂNCIAS AMOSTRAIS */
var=j(ncexp,1,.);
nrep=j(ncexp,1,.);
j=0;
do i=1 to ncexp;
  soma=0;
  somaq=0;
  controle=0;
  nrep[i] = 0;
  do while(controle=0);
    j=j+1;
    nrep[i]=nrep[i]+1;
    soma=soma+resposta[j];
    somaq=somaq+resposta[j]**2;
    if j=n_amost then do;
      controle=1;
      goto fim;
    end;
    if cexp[j+1] ^= cexp[j] then controle=1;
  fim:
  end;
  var[i] = (somaq - soma**2/nrep[i])/(nrep[i]-1);
end;
nrep2=j(ntrat,1,.);
j=0;
do i=1 to ntrat;
  controle=0;
  nrep2[i] = 0;
  do while(controle=0);
    j=j+1;
    nrep2[i]=nrep2[i]+1;
    if j=n_amost then do;
      controle=1;
      goto fim2;
    end;
    if trat[j+1] ^= trat[j] then controle=1;
  fim2:
  end;
end;

/* MODELO INICIAL DA MÉDIA - MQO */
difuso=1;                                     /*indicador para efetuar teste de falta de ajuste e ...*/
y = resposta;                                 /*construir pesos para e.q.m. e s2*/
x = xx;
RUN REG;                                      /*chama programa de regressão*/
r = resid/sqrt(1-h);                          /*correção dos resíduos*/
print 'Modelo da média - estimação por MQO';
print nomesX coef Err_Padr t probab;
print 'Teste do modelo';
print sqt sqr sqe glt glr gle;
print f_mod p_mod r2;

```

```

print 'Teste de falta de ajuste';
print sqe sqfalta sqpe gle glfalta glpe;
print f_falta p_falta;

```

/* CÁLCULO DOS RESÍDUOS QUADRÁTICOS MÉDIOS */

```

rq = r##2;
rqm=j(ncexp,1,.);
fa=0;
do i=1 to ncexp;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rq[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;

```

/* MÉTODO MISTO: RQS2 */

/* MODELO INICIAL DA VARIÂNCIA - RQS2 */

```

vrs = j(ncexp,1,.);
difuso=0;
rqv = pmod#rqm + qmod#var; /*medida que agrega e.q.m. e s2*/
rqvmed = rqv[+]/ncexp;
cte = cmed*rqvmed;
y = log(rqv+cte);
x = zz2;
RUN REG;
coef[1]=log(rqvmed); /*muda estimativa do termo constante*/
vrs = exp(x*coef);
print 'Modelo da variância - estimação por MQO';
print nomesZ coef Err_Padr t probab;
print 'Teste do modelo';
print sqt sqr sqe glt glr gie;
print f_mod p_mod r2;

```

/* PROCESSO ITERATIVO DO ESTIMADOR LOG(p.RQ + q.S2) */

DO ITER=1 TO NITER;

/* MODELOS DAS MÉDIAS USANDO MQG - RQS2*/

```

difuso=1;
va=j(n_amost,1,.); /*extende a dimensão do vetor de variâncias preditas*/
cum = 0;
do i=1 to ncexp;
  if i > 1 then cum = cum + nrep[i-1];
  ii = nrep[i];
  do j=1 to ii;
    va[cum+j] = vrs[i];
  end;
end;
w = 1/sqrt(va); /*raiz dos pesos para MQG*/
y = resposta # w; /*transf. de variáveis para MQG na rotina de MQO*/
x = xx # w;
RUN REG;
xb=xx*coef;
rnovo= (resposta-xb)/sqrt(1-h); /*novos resíduos corrigidos*/
rqnovo=rnovo##2;
if iter=niter then do;
  pred_med=yhat;
  res_med=resid;
  print 'Modelo da média - estimação por MQG iterativo';
  print nomesX coef Err_Padr t probab;
  print 'Teste do modelo';
  print sqt sqr sqe glt glr gie;
  print f_mod p_mod r2;
end;

```

```

print 'Teste de falta de ajuste';
print sqe sqfalta sqpe gle glfalta glpe;
print f_falta p_falta;
end;

/* MODELOS DA VARIÂNCIAS - RQS2*/
difuso=0;
fa=0;
do i=1 to ncexp;
  soma=0;
  do j=1 to nrep[i];
    if i=1 then jj=j;
    else jj=j+fa;
    soma=soma+rqnovo[jj];
  end;
  rqm[i]=soma/nrep[i];
  fa=fa+nrep[i];
end;
rqv = pmod#rqm + qmod#var;
y = LOG(rqv +cte);
x = zz2;
RUN REG;
coef[1]=log(rqvmed);
vrs = exp(x*coef);
if iter=niter then do;
  pred_var=yhat;
  res_var=resid;
  print 'Modelo da variância - estimação por MQG iterativo';
  print nomesX coef Err_Padr t probab;
  print 'Teste do modelo';
  print sqt sqr sqe git glr gle;
  print f_mod p_mod r2;
end;
END;
print 'Valores preditos e resíduos dos dois modelos,
      usando os pesos (1/s2) no modelo da média e transf. log no modelo da variância';
print pred_med res_med pred_var res_var;
finish ESTIMA;

run ESTIMA;

```

I - Estudo de caso nº 3

Programa computacional do exemplo de formulação de uma massa cerâmica (Bernardin, 1994). Envolve uma aplicação do algoritmo MQG - SUR desenvolvido no capítulo 4.

```

***** SERIMV.SAS *****
*** Sistema de Equações de Regressão Incluindo Modelos para as Variâncias ***
*****

* Lista das variáveis independentes X1-Xk e dependentes Y1-YL. Colocação dos dados;
* Equações das médias construídas com MQG - SUR e das variâncias com logS2 ;
* . = valores faltantes (deletados);

data dad;
  input cexp x1-x8 y1-y3;
  cards;
01 +1      -1 -1 +1 +1 -1 -1 -1 08.9 41.1 5.5
01 +1      -1 -1 +1 +1 -1 -1 -1 09.2 39.0 4.8
01 +1      -1 -1 +1 +1 -1 -1 -1 08.0 36.9 6.2
01 +1      -1 -1 +1 +1 -1 -1 -1 08.7 39.2 5.7

```

01	+1	-1	+1	+1	-1	-1	-1	-1	08.7	35.9	5.5
02	+1	-1	00	00	00	00	00	00	11.1	45.5	3.0
02	+1	-1	00	-1	00	00	00	00	11.6	55.5	2.8
02	+1	-1	00	-1	00	00	00	00	12.4	54.7	2.5
02	+1	-1	00	-1	00	00	00	00	11.6	32.0	3.0
02	+1	-1	00	-1	00	00	00	00	9.9	.	3.4
03	+1	-1	00	-1	00	-1	+1	+1	9.4	43.6	4.8
03	+1	-1	+1	00	-1	+1	+1	+1	8.5	36.2	5.5
03	+1	-1	+1	00	-1	+1	+1	+1	8.5	35.3	5.7
03	+1	-1	+1	00	-1	+1	+1	+1	6.8	37.1	6.3
03	+1	-1	+1	00	-1	+1	+1	+1	.	.	.
04	+1	00	-1	+1	00	00	+1	+1	12.6	52.7	0.9
04	+1	00	-1	+1	00	00	+1	+1	13.6	53.5	0.4
04	+1	00	-1	+1	00	00	+1	+1	11.6	47.0	1.3
04	+1	00	-1	+1	00	00	+1	+1	10.1	31.1	1.8
04	+1	00	-1	+1	00	00	+1	+1	12.1	50.9	1.1
05	+1	00	00	-1	-1	-1	-1	-1	9.4	50.0	0.8
05	+1	00	00	-1	-1	-1	-1	-1	9.9	48.3	0.6
05	+1	00	00	-1	-1	-1	-1	-1	9.6	50.1	0.6
05	+1	00	00	-1	-1	-1	-1	-1	9.2	49.9	0.7
05	+1	00	00	-1	-1	-1	-1	-1	9.4	56.2	0.5
06	+1	00	+1	00	+1	-1	00	00	9.6	35.9	6.7
06	+1	00	+1	00	+1	-1	00	00	9.4	39.7	7.0
06	+1	00	+1	00	+1	-1	00	00	8.5	27.2	7.3
06	+1	00	+1	00	+1	-1	00	00	6.8	33.6	8.0
06	+1	00	+1	00	+1	-1	00	00	8.2	.	7.5
07	+1	+1	-1	+1	+1	00	+1	00	9.9	45.1	2.3
07	+1	+1	-1	+1	+1	00	+1	00	11.1	48.7	1.5
07	+1	+1	-1	+1	+1	00	+1	00	11.1	50.9	1.7
07	+1	+1	-1	+1	+1	00	+1	00	10.6	48.2	2.0
07	+1	+1	-1	+1	+1	00	+1	00	.	.	.
08	+1	+1	00	00	-1	+1	-1	-1	6.6	31.2	9.0
08	+1	+1	00	00	-1	+1	-1	-1	6.4	25.3	10.2
08	+1	+1	00	00	-1	+1	-1	-1	5.9	22.8	10.5
08	+1	+1	00	00	-1	+1	-1	-1	5.9	27.5	10.6
08	+1	+1	00	00	-1	+1	-1	-1	6.8	31.9	9.3
09	+1	+1	+1	+1	-1	00	-1	00	7.8	32.8	6.0
09	+1	+1	+1	+1	-1	00	-1	00	8.0	40.8	5.7
09	+1	+1	+1	+1	-1	00	-1	00	8.5	36.6	5.3
09	+1	+1	+1	+1	-1	00	-1	00	8.5	46.6	5.5
09	+1	+1	+1	+1	-1	00	-1	00	8.5	.	6.7
10	-1	-1	00	-1	00	00	-1	00	13.4	60.6	0.5
10	-1	-1	00	-1	00	00	-1	00	13.4	60.0	0.5
10	-1	-1	00	-1	00	00	-1	00	13.6	68.4	0.2
10	-1	-1	00	-1	00	00	-1	00	13.4	60.8	0.7
10	-1	-1	00	-1	00	00	-1	00	12.4	51.4	1.0
11	-1	-1	00	+1	+1	+1	+1	00	9.6	41.2	3.9
11	-1	-1	00	+1	+1	+1	+1	00	10.6	53.0	4.5
11	-1	-1	00	+1	+1	+1	+1	00	8.9	37.0	3.3
11	-1	-1	00	+1	+1	+1	+1	00	7.5	30.1	3.0
11	-1	-1	00	+1	+1	+1	+1	00	8.9	41.6	3.5
12	-1	-1	+1	-1	00	-1	-1	+1	8.5	42.1	5.5
12	-1	-1	+1	-1	00	-1	-1	+1	8.7	43.4	5.9
12	-1	-1	+1	-1	00	-1	-1	+1	7.8	37.7	6.2
12	-1	-1	+1	-1	00	-1	-1	+1	.	.	.
12	-1	-1	+1	-1	00	-1	-1	+1	.	.	.
13	-1	00	-1	-1	-1	-1	+1	00	12.9	53.6	2.8
13	-1	00	-1	-1	-1	-1	+1	00	12.9	46.2	2.7
13	-1	00	-1	-1	-1	-1	+1	00	11.6	39.8	3.2
13	-1	00	-1	-1	-1	-1	+1	00	8.2	29.5	4.5
13	-1	00	-1	-1	-1	-1	+1	00	11.1	40.6	3.4
14	-1	00	00	+1	00	-1	+1	+1	7.1	35.5	6.1
14	-1	00	00	+1	00	-1	+1	+1	7.3	39.9	6.4
14	-1	00	00	+1	00	-1	+1	+1	8.2	35.2	5.0
14	-1	00	00	+1	00	-1	+1	+1	7.3	31.5	6.5
14	-1	00	00	+1	00	-1	+1	+1	5.9	.	7.0
15	-1	00	+1	+1	00	+1	00	-1	12.9	41.1	0.2
15	-1	00	+1	+1	00	+1	00	-1	12.4	39.0	0.4
15	-1	00	+1	+1	00	+1	00	-1	12.6	36.9	0.5
15	-1	00	+1	+1	00	+1	00	-1	12.6	39.2	0.4
15	-1	00	+1	+1	00	+1	00	-1	12.9	35.9	0.3

```

16 -1      +1 -1 00 00 +1 -1 00 8.2 40.8 4.4
16 -1      +1 -1 00 00 +1 -1 00 9.2 43.8 3.9
16 -1      +1 -1 00 00 +1 -1 00 9.2 48.6 4.0
16 -1      +1 -1 00 00 +1 -1 00 8.5 46.9 4.3
16 -1      +1 -1 00 00 +1 -1 00 8.7 46.2 4.1
17 -1      +1 00 +1 -1 -1 00 +1 9.9 43.4 5.6
17 -1      +1 00 +1 -1 -1 00 +1 10.1 40.9 5.0
17 -1      +1 00 +1 -1 -1 00 +1 9.6 37.9 6.1
17 -1      +1 00 +1 -1 -1 00 +1 10.1 40.6 5.2
17 -1      +1 00 +1 -1 -1 00 +1 8.9      7.0
18 -1      +1 +1 -1 +1 00 +1 -1 7.3 36.0 8.1
18 -1      +1 +1 -1 +1 00 +1 -1 6.4 30.8 9.2
18 -1      +1 +1 -1 +1 00 +1 -1 4.6 20.7 10.5
18 -1      +1 +1 -1 +1 00 +1 -1 6.4 30.1 9.4
18 -1      +1 +1 -1 +1 00 +1 -1 7.8 30.7 7.7

```

```

Data dados; /* Eliminação das observações com valor faltante em alguma variável*/
set dad;
array resp(i) y1-y3;
do over resp;
if resp=. then delete;
end;

proc iml; /*Procedimento de linguagem matricial*/
options; pagesyse=60;
/* PROGRAMA AUXILIAR COM OS DADOS INICIAIS */
START CONTROLE;
ncexp = 18; /*número de pontos experimentais (tratamentos)*/
kk = 3; /*número de variáveis dependentes*/
cmed=0.02; /*prop. da média das variâncias amostrais a ser somada na transf. log.*/
niter = 2; /*número de iterações do MQG*/
use dados;
read all;
my = y1 || y2 || y3; /* criação da matriz de var. dependentes */
n_amost = nrow(my); /*número total de observações multivariadas*/
x22 = x2##2; /*termos quadráticos e de interações*/
x33 = x3##2;
x44 = x4##2;
x55 = x5##2;
x66 = x6##2;
x77 = x7##2;
x88 = x8##2;
x12 = x1#x2;
X15 = X1#X5;
x23 = x2#x3;
x36 = x3#x6;
x47 = x4#x7;
x56 = x5#x6;
x57 = x5#x7;
x67 = x6#x7;
um = j(n_amost,1,1); /*matrizes de planejamento do modelo das médias*/
xx1 = um || x2 || x3 || x5 || x6 || x22 || x33 || x44 || x55 || x66 || x77 || x56;
xx2 = um || x2 || x3 || x5 || x6 || x77 || x23 || x36;
xx3 = um || x2 || x3 || x5 || x6 || x7 || x22 || x33 || x44 || x55 || x77 || x12 || x23 || x47;
nomexx1={cte, x2, x3, x5, x6, x22, x33, x44, x55, x66, x77, x56};
nomexx2={cte, x2, x3, x5, x6, x77, x23, x36};
nomexx3={cte, x2, x3, x5, x6, x7, x22, x33, x44, x55, x77, x12, x23, x47};
nomeX = nomexx1 || nomexx2 || nomexx3;
/*matrizes de planejamento dos modelos das variâncias*/
zz1 = um || x4 || x5 || x7 || x8 || x55 || x77 || x88;
zz2 = um || x2 || x7 || x66 || x88;
zz3 = um || x6 || x7 || x22 || x55 || x15 || x57;
nomezz1={cte, x4, x5, x7, x8, x55, x77, x88};
nomezz2={cte, x2, x7, x66, x88};
nomezz3={cte, x6, x7, x22, x55, x15, x57};
nomeZ = nomezz1 || nomezz2 || nomezz3;

```

```

pp = j(kk,1,.);
qq = j(kk,1,.);
pp[1] = ncol(xx1);
pp[2] = ncol(xx2);
pp[3] = ncol(xx3);
qq[1] = ncol(zz1);
qq[2] = ncol(zz2);
qq[3] = ncol(zz3);
finish CONTROLE;

/* PROGRAMA AUXILIAR PARA CONSTRUÇÃO DAS MATRIZES DE PLANEAMENTO DOS
MODELOS DAS VARIÂNCIAS */
start PLANVAR;
zb = j(ncexp,nparv,.);
do j=1 to nparv;
    zb[1,j] = zz[1,j];
end;
ii = 1;
do i=2 to n_amost;
    i1=i-1;
    if cexp[i] ^= cexp[i1] then do;
        ii=ii+1;
        do j = 1 to nparv;
            zb[ii,j] = zz[i,j];
        end;
    end;
end;
finish PLANVAR;

/* PROGRAMA AUXILIAR PARA REGRESSAO USANDO MQO ou MQG*/
start REG;
nobs=nrow(x);
npar=ncol(x);
xxi=inv(x`*x);
coef=xxi*x`*y;
yhat=x*coef;
resid = y-yhat;
sqe=resid`*resid;
gle=nobs-npar;
qme=sqe/gle;
covb=xxi#qme;
Err_padr=sqrt(vecdiag(covb));
t=coef/Err_padr;
probab = 1-probf(t#t, 1,gle);
if mqp = 0 then do;
    ybar = y[+]/nobs;
    desvio=y-ybar;
end;
if mqp = 1 then do;
    w = w#w;
    yp = yst#w;
    ybar = yp[+]/w[+];
    desvio=(yst-ybar)#sqrt(w);
    resid = (yst-xst*coef)#sqrt(w);
    sqe=resid`*resid;
end;
sqt=desvio`*desvio;
sqreg=sqt-sqe;
glreg=npar-1;
qmreg=sqreg/glreg;
f = qmreg/qme;
probf = 1 - probf(f,glreg,gle);
r2 = 1 - sqe/sqt;
r2aj = 1 - (1-r2)*((nobs-1)/gle);

/*definições iniciais*/
/*estimativa dos coeficientes*/
/*valores preditos*/
/*resíduos*/
/*soma de quadrados dos erros*/
/*erros padrão*/
/*estatística para o teste t de cada coeficiente*/
/*probab. sig. do teste t*/
/*MQO: desvios totais*/
/*MQG: desvios totais e resíduos*/
/*soma de quadrado dos erros na métrica do MQG*/
/*soma de quadrados total*/
/*soma de quadrados devido à regressão*/
/*teste F do modelo*/
/* coeficiente de determinação R2*/
/* R2 ajustado*/

```



```

h = j(nobs, 1,.);
do i=1 to nobs;
  h[i] = x[i,]*xxi*x[i,];
end;
if replic=1 then do;
  ymed=j(nobs, 1,.);
  fa=0;
  do i=1 to ncexp;
    soma=0;
    do j=1 to nrep[i];
      if i=1 then jj=j;
      else jj=j+fa;
      soma=soma+y[jj];
    end;
    do j=1 to nrep[i];
      if i=1 then jj=j;
      else jj=j+fa;
      ymed[jj] = soma/nrep[i];
    end;
    fa=fa+nrep[i];
  end;
  afast=y-ymed;
  sqpe=afast*afast;
  sqlof=sqe-sqpe;
  gllof=ncexp-npar;
  glpe=nobs-ncexp;
  f_falta=(sqlof/gllof)/(sqpe/glpe);
  probf_=1-probf(f_falta,gllof,glpe);
end;
Finish REG;

```

/*diagonal principal da matriz H para correção dos resíduos*/

/*somente para quando se tem replicações (mod. das médias)*/

/*teste de falta de ajuste*/

/*soma de quadrados de erro puro*/

/*soma de quadrados da falta de ajustamento*/

/* PROGRAMA AUXILIAR PARA MÍNIMOS QUADRADOS GENERALIZADOS EM SISTEMAS DE EQUAÇÕES (MQG-SUR)*/

```

start SUR;
  npar = ncol(x);
  nt = nrow(x);
  dfe = nt - npar;
  yt = y # w;
  xt = x # w;
  is = inv(ss);
  isa = is @ i(nobs);
  A = inv(xt'*isa*xt);
  bb = A * xt'*isa*yt;
  yhat = xt*bb;
  resid = yt - yhat;
  mresid = j(nobs,neq,.);
  mse = j(neq,neq,.);
  p=0;
  do k=1 to neq;
    do i=1 to nobs;
      mresid[i,k] = resid[p+i];
    end;
    p = p+nobs;
  end;
  sse = mresid*mresid;
  do k=1 to neq;
    do l=1 to neq;
      mse[k,l] = sse[k,l] / sqrt((nobs-npareq[k])*(nobs-npareq[l]));
    end;
  end;
  imse = inv(mse)@i(nobs);
  p = 0;
  q = 0;

```

/*definições iniciais*/

/*transformação para MQG (correção da heterocedasticidade)*/

/* Obtenção do vetor de parâmetros (bb) */

/* Cálculo dos valores preditos e resíduos no modelo

/* Inferência sobre os coeficientes */

```

do k=1 to neq;                                /*separação das estatísticas por equação para impressão*/
  l1 = p + 1;
  l2 = l1 + nobs - 1;
  c1 = q + 1;
  c2 = c1 + npareq[k] - 1;
  xts = xt[l1:l2,c1:c2];
  imses = imse[l1:l2, l1:l2];
  covb = inv(xts*imses*xts);
  Err_Padr = sqrt(vecdiag(covb));
  coef = bb[c1:c2];
  t = coef / Err_Padr;
  probab = 1-probf(t#t, 1, dfe);
  nomes = nome[c1:c2];
  print 'Equação' k;

  print nomes coef Err_Padr t probab;
  p = p+nobs;
  q = q + npareq[k];
end;
Finish SUR;

/* PROGRAMA AUXILIAR PARA CÁLCULO DA MATRIZ DE CORRELAÇÕES */
Start CORR;
corr = i(kk);
nobs=nrow(mat);
med = mat[+,] / nobs;
mat = mat - repeat(med,nobs,1);
do k=1 to kk-1;
  do l=k+1 to kk;
    rkl = mat[,k]#mat[,l];
    num=rkl[+];
    rk2=mat[##,k];
    rl2=mat[##,l];
    corr[k,l] = num/sqrt(rk2*rl2);
    corr[l,k] = corr[k,l];
  end;
end;
Finish CORR;

/* PROGRAMA PRINCIPAL - SERIMV */
start SERIMV;
xx = block(xx1, xx2, xx3);                    /*matrizes de planejamento ampliadas*/
zz = block(zz1b, zz2b, zz3b);
nn = n_amostr*kk;                             /*número total de obs. (todas as variáveis)*/
m_var=j(ncexp, kk,.);                         /* cálculo das variâncias amostrais e do número de replicações*/
nrep=j(ncexp, 1, 0);
do k = 1 to kk;
  j = 0;
  do i=1 to ncexp;
    soma=0;
    somaq=0;
    controle=0;
    do while(controle=0);
      j=j+1;
      if k=1 then nrep[i]=nrep[i]+1;
      soma=soma+my[j,k];
      somaq=somaq+my[j,k]**2;
      if j=n_amostr then do;
        controle=1;
        goto fim;
      end;
      if cexp[j+1] ^= cexp[j] then controle=1;
    fim;
  end;
  m_var[i,k] = (somaq - soma**2/nrep[i])/(nrep[i]-1);

```

```

end;
end;
print 'Variâncias amostrais';
print m_var;

/* MODELOS UNIVARIADOS PARA AS MÉDIAS */

replic=1;
mqp=0;
p=0;
q=0;
r = j(n_amost,kk,.);
predito = j(n_amost,kk,.);
do k=1 to kk;
  l1 = p+1;
  l2 = l1 + n_amost-1;
  c1 = q+1;
  c2 = c1 + pp[k]-1;
  y = my[,k];
  x = xx[l1:l2 , c1:c2];
  RUN REG;
  print 'Modelo univariado para as médias - Equação:' k;
  nomes = nomeX[c1:c2];
  print nomes coef Err_Padr t probab;
  print 'ANOVA do modelo:' f greg gle probf;
  print 'Coeficiente de determinação:' R2 R2aj;
  print 'Teste de falta de ajuste:' f_falta gllof glpe probf_;
  predito[,k] = yhat; /*gráfico dos resíduos*/
  r[,k] = resid/sqrt(1-h); /*correção dos resíduos*/
  mr = yhat || (r[,k]/sqrt(qme));
  call pgraf(mr, '*', 'valor predito', 'resíduo studentizado', 'Gráfico dos resíduos');
  p = p+n_amost;
  q = q + pp[k];
end;

/* MODELOS UNIVARIADOS PARA AS VARIÂNCIAS, COM TRANSF. LOG. */

replic=0;
p=0;
q=0;
v=j(ncexp,kk,.);
resvar = j(ncexp,kk,.);
cte=j(kk,1,.);
cte3=j(kk,1,.);
do k=1 to kk; /*análise por equação*/
  varmed = m_var[,k]/ncexp; /*variância média para transf. log. modificada*/
  cte[k] = cmed*varmed; /*constante para transf. log. modificada*/
  cte3[k] = 0.005*varmed; /*constante para controle das variâncias*/
preditas*/
  l1 = p+1;
  l2 = l1 + ncexp-1;
  c1 = q+1;
  c2 = c1 + qq[k]-1;
  y = log(m_var[,k] + cte[k]); /*transformação log. modificada*/
  x = zz[l1:l2 , c1:c2];
  RUN REG;
  v[,k] = exp(yhat) - cte[k]; /*transf. inversa (predição das variâncias)*/
  coef[1]=log(varmed); /*Estimativa do termo constante: média das variâncias*/
  nomes = nomeZ[c1:c2];
  print 'Modelo univariado para as variâncias - Equação:' k;
  print nomes coef Err_Padr t probab;
  print 'ANOVA do modelo:' f greg gle probf;
  print 'Coeficiente de determinação:' R2 R2aj;
  p = p+ncexp;
  q = q + qq[k];
  do i=1 to ncexp; /*limite mínimo para as variâncias preditas*/
    if v[i,k] < cte3[k] then v[i,k]=cte3[k];
  end;
end;

```

```

resvar[,k]=resid;                                /*gráfico dos resíduos*/
resid = resid / sqrt(qme*(1-h));
vr = yhat || resid;
call pgraf(vr, '*', 'valor predito', 'resíduo studentizado', 'Gráfico dos resíduos');
end;

/* EXPANSÃO DA MATRIZ DE VARIÂNCIAS PREDITAS */

va=j(n_amost,kk,.);
do k=1 to kk;
  cum = 0;
  do i=1 to ncexp;
    if i > 1 then cum = cum + nrep[i-1];
    ii = nrep[i];
    do j=1 to ii;
      va[cum+j,k] = v[i,k];
    end;
  end;
end;

/* MODELOS UNIVARIADOS PARA AS MÉDIAS USANDO MÍNIMOS QUADRADOS
GENERALIZADOS */
replic=1;
rhomo = j(n_amost,kk,.);
p=0;
q=0;
mqp=1;
do k=1 to kk;
  l1 = p+1;
  l2 = l1 + n_amost-1;
  c1 = q+1;
  c2 = c1 + pp[k]-1;
  yst = my[,k];
  xst = xx[l1:l2, c1:c2];
  w = 1/sqrt(va[,k]);                                /*pesos para o MQG*/
  y = my[,k] # w;
  x = xx[l1:l2, c1:c2] # w;
  RUN REG;
  nomes = nomeX[c1:c2];
  print 'Modelo univariado para as médias usando MQG - Equação:' k;
  print nomes coef Err_Padr t probab;
  print 'ANOVA do modelo:' f gireg gle probf;
  print 'Coeficiente de determinação:' R2 R2aj;
  print 'Teste de falta de ajuste:' f_falta gllf glpe probf_;
  if k=1 then bb = coef; else bb = bb // coef;
  rhomo[,k] = resid/sqrt(1-h);                        /*resíduos corrigidos na métrica do MQG*/
  p = p+n_amost;                                      /*gráfico dos resíduos*/
  q = q + pp[k];
  mr = yhat || (rhomo[,k]/sqrt(qme));
  call pgraf(mr, '*', 'valor predito', 'resíduo studentizado', 'Gráfico dos resíduos');
end;

/* PREPARAÇÃO DO PROCESSO MULTIVARIADO */
/* cálculo da matriz de correlações*/
mat = rhomo;
RUN CORR;
p = 0;
do k=1 to kk;
  if k=1 then yy = my[,1]; else yy = yy // my[,k];
end;

do iter=1 to niter;
  print ***** ITERAÇÃO NÚMERO:' iter '*****;
/* ESTIMATIVA MULTIVARIADA DAS MEDIAS, USANDO O PROGRAMA AUXILIAR SUR */
  print 'MODELO MULTIVARIADO PARA AS MÉDIAS - MQG';
  vv = j(nn,1,.);
  do k=1 to kk;
    if k=1 then vv=va[,1]; else vv = vv // va[,k];
  end;
  nobs=n_amost;

```

```

x = xx;                                /*matrizes e vetores de todas as variáveis*/
y = yy;
nome = nomeX;
w = 1/sqrt(vv);                        /*raiz dos pesos para transformar x e y*/
npareq = pp;                           /*número total de parâmetros*/
ss = corr;                             /*matriz de correlações*/
neq = kk;                              /*número de equações*/
RUN SUR;

/* CALCULANDO OS RESÍDUOS HOMOCEDÁSTICOS (na métrica do MQG)*/
hh = j(nn,1,.);
mm = xx*bb;
mmt=xt*bb;
rr = yy - mm;
rrhomo = yt - mmt;
do i=1 to nn;                          /*correção dos resíduos*/
    hh[i] = xt[i,]*A*xt[i,];
    rr[i] = rr[i]/sqrt(1-hh[i]);
    rrhomo[i] = rrhomo[i]/sqrt(1-hh[i]);
end;
p = 0;
do k=1 to kk;                          /*transf. do vetor de resíduos em matriz por equação*/
    do i=1 to n_amost;
        rhomo[i,k] = rrhomo[p+i];
    end;
    p = p+n_amost;
end;
mat = rhomo;                          /*cálculo das correlações com os novos resíduos*/
run corr;
print 'Matriz de correlações dos resíduos homocedásticos';
print corr;
END;
/* APRESENTAÇÃO DAS MÉDIAS, VARIÂNCIAS E RESÍDUOS EM FORMAS MATRICIAIS*/
q = 0;
m=j(n_amost,kk,.);
do k=1 to kk;
    do i=1 to n_amost;
        m[i,k] = mm[q+i];            /* Estimativa das médias */
    end;
    q = q+n_amost;
end;
print 'ESTIMATIVA FINAL DAS MÉDIAS';
print m;
print 'ESTIMATIVA FINAL DAS VARIÂNCIAS';
print v;
finish SERIMV;

/* PROGRAMA DE EXECUÇÃO */
start PROG;
run CONTROLE;                        /*para ler os dados e as características iniciais*/
zz=zz1;                             /*ajusta a dimensão da matriz de planejamento do modelo da variância*/
npurv=ncol(zz);                      /*processo repetido para cada equação*/
run PLANVAR;
zz1b = zb;
zz=zz2;
npurv=ncol(zz);
run PLANVAR;
zz2b = zb;
zz=zz3;
npurv=ncol(zz);
run PLANVAR;
zz3b = zb;
run SERIMV;                          /*chama o programa principal*/
finish PROG;
run PROG;

```